Engineering of secondary metabolite production in streptomycetes

Robertsen, Helene Lunde

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Technical University
of Denmark

Engineering of secondary metabolite production in streptomycetes

Ph.D. Thesis

Helene Lunde Robertsen

November 2017

DTU Biosustain
The Novo Nordisk Foundation Center for Biosustainability

Education is what remains after one has forgotten what one has learned in school

Albert Einstein

Preface

This thesis is written as a partial fulfilment of the requirements to obtain a PhD degree at the Technical University of Denmark. The work presented in this dissertation is the result of three years work from December 2014 – November 2017 at the Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark. The work was conducted under the supervision of supervisor Dr. Tilmann Weber and co-supervisor Professor Lone Gram, both from the Technical University of Denmark. Funding was provided by the Novo Nordisk Foundation.

The thesis was evaluated by Professor Thomas Ostenfeld Larsen, Technical University of Denmark, by Dr. Stefano Donadio, Chief Executive Officer, NAICONS Srl, Milano, Italy, and by PD Dr. Evi Stegmann, University of Tübingen.

Acknowledgment

I owe many people a great amount of appreciation for their patience, guidance, and help during the works of this PhD. First, I wish to thank scientific director Prof. Dr. Sang Yup Lee for the opportunity to work in his group. I also would like to thank my supervisor Dr. Tilmann Weber for the many fruitful discussions, for his guidance, and most of all for giving me the opportunity to pursue a career in academia.

Thanks to Dr. Ling Ding for her collaboration and help with all the analytical analyses and to Dr. Yaojun Tong for introducing me to the CRISPR-Cas9 technique. And thanks to all of my lovely co-workers! You made those long working hours a lot easier to get through.

During the last three years, I have also had the pleasure of forming many great, new friendships. During a 2,5-months stay in the group of Prof. Dr. Timo Niedermeyer in the University of Tübingen, I was introduced to the two very charming Ph.D. students Paul and Nico! I hope one day to be able to repay your hospitality and kindness.

On a personal note, a very special thanks goes to my parents. Their help and support allowed me to get through a very chaotic time in my life. I will be forever grateful for your pride and faith in me.

Finally, my biggest and most heart-warming thanks goes to my friend and co-worker postdoc Dr. Ewa Maria Musiol-Kroll, who I had the pleasure of working in close collaboration with for the first two years of the thesis. You opened my eyes to the joys and hassle of working with streptomycetes. Thank you for your huge support, your magnificent guidance in lab, and, last but not least, our always nice morning talks.

Summary

Streptomycetes are known for their ability to produce a range of different secondary metabolites, including antibiotics, immunosuppressive, anti-fungals, and anti-cancer compounds. Of these compounds, antibiotics play an important role in the clinics for treatment of both mild and severe bacterial infections. However, with the rise of multi-resistant pathogens, the demand for new antibiotics or derivatives of old ones, with improved properties, is now higher than ever.

Recent efforts in genome sequencing and mining have revealed a so far untapped potential of streptomycetes and related actinomycetes as evident from so-called "silent" biosynthetic gene clusters, whose products remain undetectable under standard laboratory conditions. These clusters harbour all information necessary for production of potentially novel bioactive compounds, and hence provide high priority candidates for engineering to activate their production. With this knowledge, the need for better molecular tools to harness the potential of the gifted microorganisms is now greater than ever. One such molecular tool, which has truly revolutionised the field of genome engineering, is the CRISPR-Cas9 genome engineering system.

In this thesis, the CRISPR-Cas9 system for genome engineering of actinomycetes was expanded for future applications in a high-throughput semi-automatic setting. First, a toolbox and workflow for construction of CRISPR plasmids, for a range of different engineering purposes was developed, including the computational prediction of suitable 20 bp protospacers for the single guide RNAs and a USER-cloning method for construction of the CRISPR plasmids. Additional improvement to the system was achieved through the development of an optimised USER assembly workflow for cheaper and faster plasmid construction. The workflow was verified by manual knock-down of two biosynthetic gene clusters in model organism *Streptomyces coelicolor* A3(2), which confirmed the applicability of the system.

A second part of the thesis was devoted to engineering of *Streptomyces collinus* Tü 365, which is a known producer of the narrow-spectrum antibiotic kirromycin.
While there exists several studies addressing the PKS scaffold biosynthesis of kirromycin, knowledge about the supply of the precursor ethylmalonyl-CoA and most of the tailoring reactions remained scarce. In this thesis, the role of the gene *kirN*,

believed to be involved in precursor supply, and the six genes *kirM, kirHIV, kirHV, kirHVI, kirOI* and *kirOII*, all predicted to be involved in tailoring reactions, were investigated by gene inactivations, complementations, and characterisation of the biosynthetic products of the generated mutants. Within our studies, four novel kirromycin derivatives were generated and characterised. Our investigations allowed for closing some of the missing gaps in the biosynthesis of kirromycin, along with providing us with a toolbox of new mutants, which produce derivatives of the original compound. These derivatives could serve as scaffolds for future bioderivatization efforts.

This thesis lays the groundwork for future engineering of streptomycetes to improve secondary metabolite production. For the USER-CRISPR-Cas9 platform, the next logical step will be to implement the workflow in a robotic setting. Furthermore, the mutants of *S. collinus* Tü 365 will be included in a derivatization platform to produce new kirromycin analogues with improved pharmacokinetic properties.

Sammenfatning

Streptomyceter er kendt for deres evne til at fremstille forskellige bioaktive stoffer, deriblandt antibiotika, immunsuppresive, anti-svampe, og anti-cancer stoffer. Særligt antibiotika har en vigtig klinisk rolle, idet de har gjort det muligt at behandle både simple og mere seriøse bakterielle infektioner hurtigt og effektivt og med minimale bivirkninger. I de seneste år er efterspørgsel på nye og forbedrede antibiotika i mellemtiden steget, idet fremdriften af multiresistente bakterier vil komme til at udgøre en seriøs trussel for fremtidens samfund.
Heldigvis har genomsekventering af disse talentfulde bakterier afsløret flere såkaldte kryptiske eller tavse gen clusters, der besidder potentialet for fremstilling af nye klinisk relevante stoffer, men forbliver udetektérbare under normale vækstforhold. En måde at udnytte dette skjulte potentiale og aktivere produktionen af de potentielt nye bioaktive stoffer er gennem brugen af molekylære værktøjer, såsom det nyudviklede CRISPR-Cas9 værktøj, til at ændre det kodende materiale i bakterierne.

Denne afhandling søger at lægge et fundament for fremtidig engineering af streptomyceter med henblik på at fremme deres produktion af klinisk relevante bioaktive stoffer. Vores tilgang til dette kan deles op i to hovedprojekter.
I første del af afhandlingen præsenteres udviklingen og optimeringen af en USER-kloningsmetode til fremstilling af plasmider til CRISPR-Cas9 genome engineering. Kloningssystemet blev søgt optimeret for at fremme metodens brug i fremtidige high-throughput, semi-automatisk systemer. Anvendeligheden af platformen blev testet manuelt i to proof-of-concept studier, hvor produktion af to bioaktive stoffer fra model-organismen *Streptomyces coelicolor* A3(2) blev hæmmet som følge af brugen af det nye plasmid-system. Med de lovende resultater fra den manuelle test af system bliver det næste store skridt at implementere kloningsmetoden på en robot og dermed teste anvendeligheden af hele arbejdsprocessen i en high-throughput baggrund.

Anden del af afhandlingen omhandler genetiske studier og engineering af *Streptomyces collinus* Tü 365, som er kendt for at fremstille antibiotikummet kirromycin. Grundet uønskede bivirkninger kan kirromycin på nuværende tidspunkt ikke bruges til behandling i klinikken. Dette gør stoffet til en ideel kandidat for bioderivatisering, idet ændringer i den molekylære sammensætning kan ændre et stofs virkningsmekanisme drastisk.

Adskillige studier i biosyntesen af kirromycin har ført til detaljeret viden om systemet, men til dato mangler der stadig information om precursor forsyning og de sidste tailoring reaktioner, der sammen fører til det endelig molekyle. Vores arbejde med syv gener, hvoraf ét var formodet involveret i precursor forsyning og seks i tailorings, har gjort det muligt for os at udfylde nogle af de manglende huller i biosyntesen af kirromycin. Ydermere resulterede vores studier i en samling af fire nye genetiske mutanter, der producerer nye kirromycin derivater, og en mutant med lavere produktion af kirromycin. Særligt den sidste mutant har høj interesse idet den kan indgå i en bioderivatiserings-platform til fremstilling af nye analoge stoffer af kirromycin med forbedrede egenskaber.

Outline of the Ph.D. thesis

The following Ph.D. thesis with the title "Engineering of secondary metabolite production in streptomycetes" contains four manuscripts organised in three main sections. Besides the manuscripts, an introduction and discussion, which both sum up the general themes of the thesis, are included in the dissertation.

After a brief introduction to metabolic engineering, the review (**manuscript I**) with the title "Towards Systems Metabolic Engineering of Streptomycetes", which was recently published in Biotechnology Journal, is presented.

Manuscript II contains the method "CRISPR-Cas9 Toolkit for Actinomycete Genome Editing", in which the overall workflow developed for genome-to-knockout/down of actinomycetes is presented, which was recently published as a chapter in the book entitled "Synthetic Metabolic Pathways".

Following this, **manuscript III** entitled "An improved USER-CRISPR-Cas9 platform for high-throughput genome editing in streptomycetes" is presented. The draft manuscript describes an optimised USER-cloning system developed for fast and efficient construction of CRISPR-(d)Cas9 constructs, which should serve as a platform to improve our future efforts in semi-automatic, high-throughput gene and gene cluster cloning of streptomycetes.

A case in which the CRISPR-Cas9 technology proved unsuitable for genome editing appeared during our work on kirromycin biosynthesis in *Streptomyces collinus* Tü 365. Instead, an alternative method was used for constructing the genetic mutants in this strain and the results from this work are presented in **manuscript IV** with the title "Filling the Gaps in the Kirromycin Biosynthesis: Deciphering the Role of Genes Involved in Ethylmalonyl-CoA Supply and Tailoring Reactions".

Table of Contents

Introduction

Biology of streptomycetes

Streptomycetes are gram-positive, soil-dwelling bacteria, which belong to the order of Actinomycetales. Actinomycetes bear morphological resemblance to fungi as evident from the formation of mycelia, which is the result of growth by tip extension and branching of the hyphae. Under sufficient growth conditions, streptomycetes form a dense vegetative mycelium[1], in which cell division occurs by cross-wall separation of the hyphae into individual compartments[2]. As a result, each compartment can contain multiple copies of the chromosome, making streptomycetes multicellular organisms[3]. When encountering nutrient depletion or other stress signals, the hyphae develop into aerial hyphae, which due to their hydrophobic surface can rise into the air and further segregate into mature spores[4]. The dormant spores can readily spread in the environment and ensure the survival of the bacteria. Once vegetative growth conditions are encountered again germination of the spores are initiated[5,6] (**Fig. 1**).

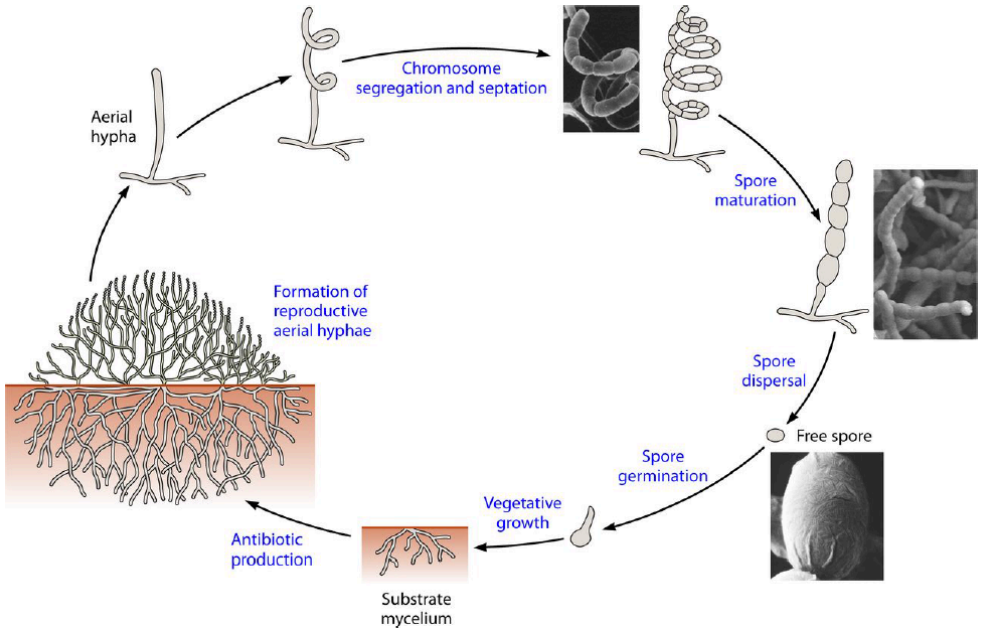


Figure 1 Schematic overview of the life cycle of streptomycetes. From Barka *et al.*[6]

Another interesting feature of streptomycetes is their abundance and importance in soil. Accounting for 95 % of Actinomycetales isolates from soil, streptomycetes produce a range of different exoenzymes, which are used to break down polymers to release carbon[7]. In addition, streptomycetes harbour the potential to produce a multitude of bioactive secondary metabolites, including, but not limited to,

antibiotics, immunosuppressants, antifungals, and anti-cancer compounds[8]. In fact, in the so-called "golden era" of antibiotic discovery, ranging from the 1940s until the early 1950s, 70 to 80 % of all newly identified natural antibiotics were isolated from *Streptomyces* species[9]. These clinically relevant compounds are produced from biosynthetic gene clusters (BGCs) encoded in the linear chromosomes of the bacteria as a part of their secondary metabolism. Although not a part of primary metabolism, the biosynthesis of bioactive secondary metabolites is believed to be of great importance to cell survival as means of self-protection in response to external stimuli, such as nutrient depletion or when faced with a competitive niche[6].

Since the discovery of several "cryptic" gene clusters in the model strain *Streptomyces coelicolor* A3(2)[10], the field saw the need for understanding the biosynthetic logic behind secondary metabolism. As a result, acquisition of high-quality data from the large, G+C-rich genomes of streptomycetes is now becoming a routine practice in many laboratories as a means of uncovering the hidden potential of the strains through genome mining[11–13]. For additional information on this topic, the reader is referred to **manuscript I** of this thesis.

Biosynthesis of secondary metabolites in streptomycetes

The genes required for biosynthesis of secondary metabolites are often clustered in the genomes of streptomycetes, a feature which has enabled the discovery of several natural products using genome mining tools such as antiSMASH[14] or PRISM[15]. The overall principle behind biosynthesis from BGCs can be divided into three main phases: basic building block supply, assembly of the building blocks, and tailoring of the precursor molecule to yield the final product[16]. For many of the complex secondary metabolites, the building blocks, which are supplied from primary metabolism or through feedings, are assembled by enzymatic machineries encoding all information required for synthesis, transport, and resistance. The genetic information stored in these mega-enzymatic complexes is often organised by an assembly line principle, in which each gene encodes modules that are further divided into separate enzymatic domains. Here, the gathered enzymatic actions of the domains in the modules are responsible for the activation and attachment of one extender unit. Complexity in biosynthesis of secondary metabolites is achieved through variations in

the three biosynthetic phases, including variation in and linkage of the building blocks and in the type of tailoring reactions encoded in the BGC.

Polyketides

The biosynthesis of polyketides is facilitated by polyketide synthases (PKSs). Polyketides are composed of simple carboxylic acid monomers, which are assembled in a manner similar to that of fatty acid (FA) biosynthesis. Assembly of the activated acyl-CoA monomers is facilitated by iterative decarboxylative Claisen-like condensation. The polyketide precursor can then undergo a series of reductive steps, which is facilitated by ketoreductases (KRs), dehydratases (DHs), and enoyl reductases (ERs). In contrast to FA biosynthesis, the degree of reduction is subject to variation in PKSs, a feature which gives rise to higher complexity of the polyketides. Upon assembly, the precursor molecule is released from the PKS by, most commonly, a thioesterase (TE), at which point the molecule can be subjected to additional enzymatic tailoring reactions, such as methylations, hydroxylations, halogenations, and ring formations.

Depending on their domain architecture, PKSs are further grouped into three different types (type I-III PKSs), that can be either modular (type I) or iterative (types I, II, III), which determines whether the domains are used in a single or repetitive fashion, respectively.

Type I PKSs are the most complicated and versatile of the three PKS classes[17] and are divided into the iterative and modular type. The simplest structure is that of the iterative type I PKS, in which one module is used repeatedly for chain elongation. The presence of iterative type I PKSs is relatively rare in prokaryotes and only gives rise to small, aromatic compounds such as the "warhead" structures of enediynes anticancer antibiotics[18,19].

In contrast to the iterative type I PKSs, modular type I PKS can give rise to larger and more complex polyketides such as the antibiotics erythromycin and pikromycin or the anti-cancer compound chartreusin. The "minimal PKS module", which is responsible for the chain elongation with one extender unit at a time, is composed of three essential domains, namely the acyltransferase (AT), ketosynthase (KS) and the acyl carrier protein (ACP). During each round of chain elongation, the AT selects and

loads the extender unit onto the ACP from which the formation of C-C-bond, catalysed by the KS, takes place. The optional presence of β-keto-processing enzymes, KR, DH, and ER, determines the level of reduction in each module. Recently, the use of cryogenic electron microscopy (cryo-EM) on the pikromycin type I PKS module five, encoded by *pikAIII*, has helped to understand the organisational changes occurring during assembly[20,21] (**Fig. 2**). During chain elongation, the ACP, carrying the intermediate chain, is moved between the enzymatic domains to interact with these and finally is ejected out to interact with the sixth module in the pikromycin assembly line.

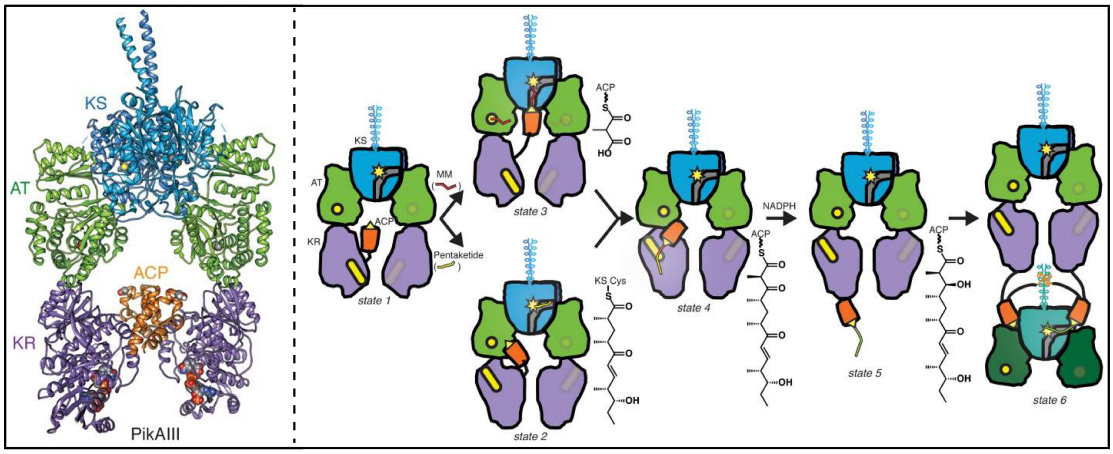


Figure 2 Cryo-EM of pikromycin PKS module five (holo-PikAIII) (left figure)[21] and schematic overview of substrate processing of the same module PikAIII (right figure)[20]. During the transition from state 1 to 2, the KR (purple) undergoes an end-to-end flip, which places the ACP (orange) close to the AT (green) domain. Upon loading of methylmalonyl (MM, red line), the ACP is moved to the bottom entrance of the KS (blue) (state 3). The extension of the pentaketide (yellow and red lines) moves the precursor bound to the ACP to the KR domain (state 4). In the final two states, the ACP is ejected out of the central pocket of the complex (state 5) to interact with the next module PikAIV in the assembly line (state 6). ACP, acyl carrier protein; AT, acyltransferase; KR, ketoreductase; KS, ketosynthase; MM, methylmalonyl; NADPH, Nicotinamide adenine dinucleotide phosphate.

The modular organisation of type I PKSs often follows the rule of "colinearity", which is the result of one-to-one correspondence of PKS architecture and metabolite structure. This feature has been successfully applied for bioinformatic predictions of molecular structures based on BGC architecture or vice versa[19]. However, the presence of duplicated or inactive domains complicates discovery efforts[22], as in the case of rapamycin biosynthesis in *Streptomyces hygroscopicus*. In the rapamycin BGC, four DH domains were identified as inactive based on deletions in the characteristic active site sequence, however, further sequence analysis failed to provide any explanation for the missing activities of two additional DH domains, of

two KR domains, and of one ER domain, located in two separate modules (module 3 and 6)[23,24].

Furthermore, the presence of so-called *trans*-AT PKSs, which, unlike *cis*-AT PKSs, use freestanding AT enzymes, which has the ability to load extender units to several different modules in the same type I PKS assembly line, can bring extra complexity to the molecule, which is the case for the antibiotic kirromycin[25–27].

Type II and III PKSs synthesise mostly aromatic molecules, albeit using different enzymatic approaches[17]. Type II PKSs are exclusively found in bacteria, in which they iteratively synthesise aromatic polyketides[19], such as the antibiotic actinorhodin[28] and the anti-cancer compound doxorubicin[29]. For these PKSs, the minimal set of enzymes needed for assembly is restricted to two KSs, KS_α and KS_β, and the ACP. During chain elongation, the thioester is tethered to the ACP while the $KS_\alpha KS_\beta$-complex is responsible for elongation of the chain with malonyl-CoA units exclusively. Based on the current working hypothesis, an additional enzymes should be included as a component of the minimal PKS, namely the malonyl-CoA:ACP transferase (MCAT), which is believed to recruit the malonate extender units from primary metabolism. Since most clusters of type II PKSs do not encode any MCAT, the homologous enzyme from FA biosynthesis has been speculated to be capable of interacting with the PKS by loading the malonate units to the ACP[30]. However, contradictory examples exist to this hypothesis as exemplified by the well-studied biosynthetic pathway of actinorhodin in *Streptomyces coelicolor*, in which the chemically synthesised variant of ACP was capable of undergoing self-malynation in the absence of any MCAT activity[31].

To avoid spontaneous cyclisation of the polyketide intermediate upon release from the enzymatic complex, additional cyclases (CYCs) and KRs are often encoded in the BGCs of type II PKSs to act as chaperons and reductive agents, respectively, in order to facilitate the correct cyclisation. In addition, the presence of aromatases (AROs) results in the biosynthesis of the aromatic ring systems.

Type III PKSs distinguish themselves from other PKSs by their lack of ACPs and instead the presence of a single KS-like enzyme responsible for condensation of acetate units to CoA-derivatized starter units[17]. Within the same PKS active site cavity, the linear chain is then further processed through intramolecular condensation and aromatisation. Diversity is added to the polyketide chain through variations in

choice of acyl-CoA starter units, the number of elongation rounds, and the mechanism of folding of the final molecule[17,19].

Although previously thought to be exclusive to plants, type III PKSs can also be found in bacteria[32,33]. Examples of bacterial type III PKSs involved in biosynthesis are found in the biosynthetic pathway of balhimycin in *Amycolatopsis bahlimycina*[32] and for the 1,3,6,8-tetrahydroxynaphthalene (THN) precursor, which can undergo spontaneous oxidation to yield the pigment flaviolin found in many streptomycetes[33,34]. For balhimycin biosynthesis, the type III PKS is encoded by the DpgA enzyme, responsible for provision of the non-proteinogenic amino acid 3,5-dihydroxy-phenylglycine, which is essential for biosynthesis of the antibiotic[32].

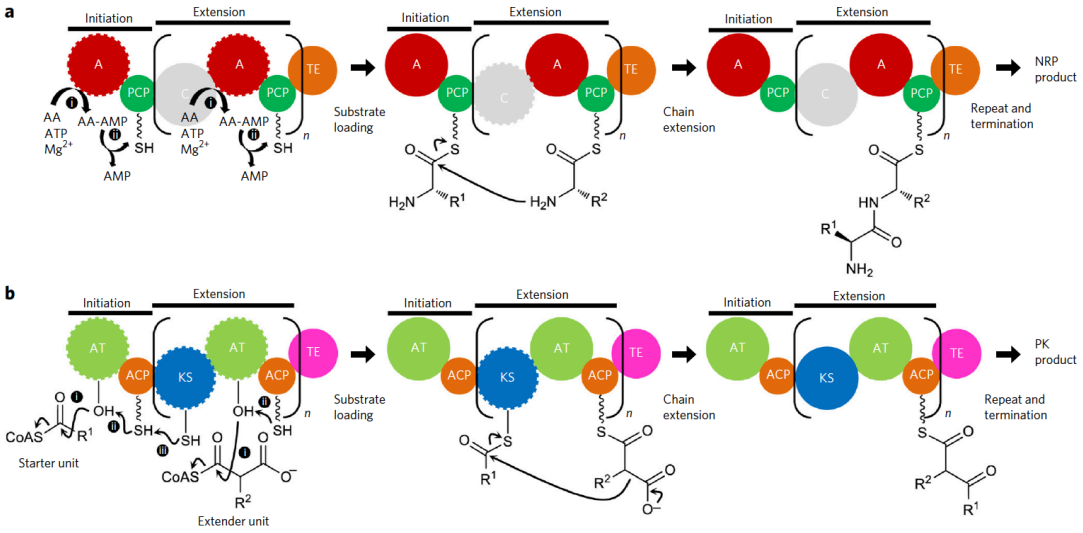


Figure 3 Schematic representations of the biosynthetic cycles of a minimal NRPS (a) and PKS (b) assembly line, depicting both initiation and extension modules. A, adenylation domain; AA, amino acid; ACP, acyl carrier protein; AT, acyltransferase; C, condensation domain; NRP, non-ribosomal peptide; KS, ketosynthase; PCP, peptidyl carrier protein; PK, polyketide; TE, thioesterase domain. From Weissman[22].

Non-ribosomally synthesised peptides

Another important group of secondary metabolites are the peptide-derived compounds, which are composed of proteogenic or non-proteogenic amino acids, assembled by non-ribosomal peptide synthetases (NRPSs) (**Fig. 3**). For NRPSs, selection and loading of amino acid extender units are catalysed by the adenylation (A) domain, which also activates the bound amino acid at the expense of ATP. The activated amino acid is in turn tethered to the peptidyl carrier protein (PCP) through a thiolation reaction resulting in an amino-acid-PCP-thioester. Due to this reaction, the PCP domain is also referred to as the thiolation (T) domain. Peptide bond formation

and further chain elongation are catalysed by the condensation (C) domain. Two common modification reactions can occur in NRPS assembly lines. The first is epimerisation (E) to convert the L-aminoacyl and L-peptidyl thioesters into their corresponding D-forms before being transferred to the next module. The second type of common modification arises from the actions of methyltransferase (MT) domains, responsible for N-methylation of the NRPS intermediate[35]. Finally, in bacteria the TE domain is most often encoded in the last module of the megaenzyme and is responsible for chain termination and release from the NRPS assembly line.

Recently, structural insight on the distinct conformations of the two NRPS modules AB3403 from the pathogen *Acinetobacter baumannii* and EntF from *Escherichia coli* has helped to elucidate the dynamics behind two essential catalytic states in the NRPS cycle[36]. More specifically, during the four states in the catalytic cycle, the module undertakes three distinct conformations, in which the initial amino acid adenylation (AA-AMP) (state I) is followed by delivery of the AA-AMP to the A domain for thioester-formation to load the PCP (state II). Upon activation, the NRPS complex converts back to the initial conformation I to deliver the PCP to the C domain to receive the upstream peptide (state III). In the final conformation (state IV) the peptide is delivered to a downstream C, TE, or reductase domain (**Fig. 4**)[36].

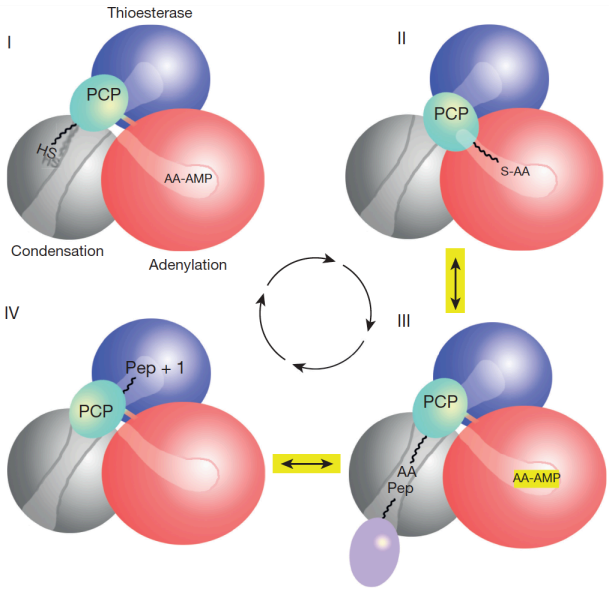


Figure 4 The NRPS catalytic cycle involving four states (I – IV)[36]. In state (I), the incoming amino acid (AA) is activated by the adenylation (A) domain. During (II), the peptidyl carrier protein (PCP) moves to the A domain resulting in thioester bond formation between co-factor and AA. In (III), the activated AA bound to PCP is moved to the condensation (C) domain at which point it can interact with an upstream peptide. Finally, in state (IV) the peptide chain is moved to a new C domain or released by a thioesterase (TE) or reductase domain. AA, amino acid; PCP, peptidyl carrier protein; Pep, peptide.

Hybrid PKS/NRPSs

As a result of the similarity between their biosynthetic machineries (**Fig. 3**), PKSs and NRPSs can collaborate to assemble hybrid polyketide-peptide metabolites[22]. Excellent examples exist of hybrid PKS/NRPS assembly lines, including that of the immunosuppressant ascomycin (FK520), which is composed of 11 PKS modules distributed on the three proteins FkbB, FkbC and FbkA, followed by one NRPS module on the protein FkbP[37]. The opposite situation also exists where polyketide units are incorporated in NRPS-derived peptides, as exemplified by the biosynthesis of the anti-cancer compound bleomycin. For bleomycin biosynthesis, the hybrid assembly line is composed of ten NRPS modules and one PKS module, which result in the incorporation of a single malonyl-CoA unit in the molecule[38]. In both of the aforementioned cases, what remains to be studied in these hybrid assembly lines is how translocation of the growing chain between the PKS/NRPS and NRPS/PKS interfaces is facilitated[39].

Kirromycin

The antibiotic kirromycin (**Fig. 5**) is an excellent example of how complex Nature's own biosynthetic machinery can be. Kirromycin was first isolated from *Streptomyces collinus* Tü 365 in 1972 by Wolf and Zähner in Tübingen, Germany[40].

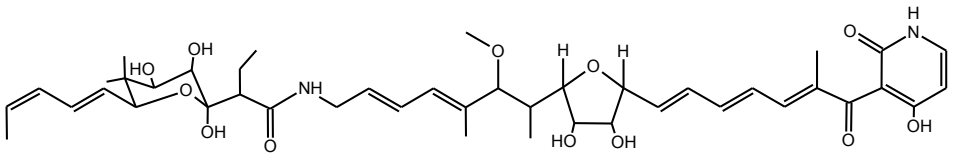


Figure 5 Structure of kirromycin.

Kirromycin is a narrow-spectrum antibiotic with activity against gram positive and gram negative bacteria, including *Enterococcus faecium*, *E. faecalis*, *Haemophilus influenzae* type B, *Neisseria gonorrhoeae*, and the malaria parasite *Plasmodium falciparum*[41–43]. The molecule belongs to the class of elfamycin antibiotics, which are grouped based on their ability to interact with prokaryote elongation factor (EF) Tu[44]. Interestingly, this group of antibiotics shows little toxicity against the mitochondrial elongation factor in eukaryotes[45]. Kirromycin binds to the interface of domains I and III of the EF Tu in complex with either GDP or GTP (**Fig. 6**), although with highest affinity for the latter[46,47]. The binding induce the EF Tu to maintain the conformation

it would adopt when bound to GTP, which then prevents release of the complex from the ribosome and ultimately results in the stalling of protein translation.

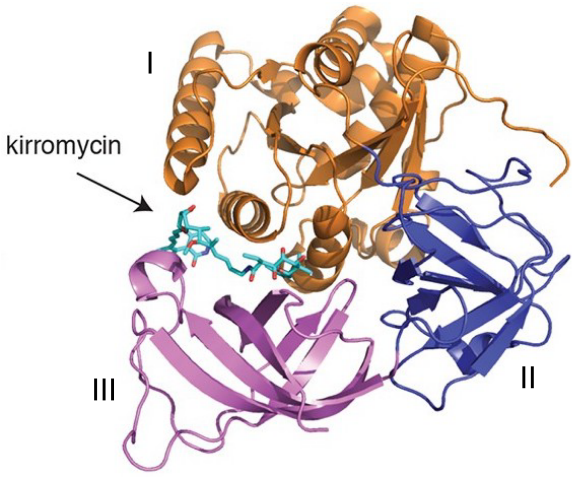


Figure 6 Crystal structure of elongation factor (EF) Tu. Kirromycin binds between domains I and III in the complex, which has been marked in the illustration. Adapted from Carelli *et al.*[48]

Kirromycin suffers from poor pharmacokinetics, which has limited its use in the clinics. However, with increased reports of resistant isolates of *N. gonorrhoeae* arising in the clinic, the group of elfamycin antibiotics might experience revitalisation in future years. In this sense, kirromycin with a reported minimal inhibitory concentration (MIC) of 0.06 µg/mL might be subject to new studies to change properties of the molecule and improve its pharmacokinetics[41].

In addition to the studies on the mode of action and clinical value of kirromycin, the molecule has also been subject to detailed biosynthetic studies. As a result, the full BGC of kirromycin was characterised in 2008, revealing an 82 kb DNA region composed of 28 genes involved in synthesis and transport (**Fig. 7**). Furthermore, genome mining of *S. collinus* Tü 365 later revealed the presence of two copies of the kirromycin BGC on the arms of the chromosome[49,50].

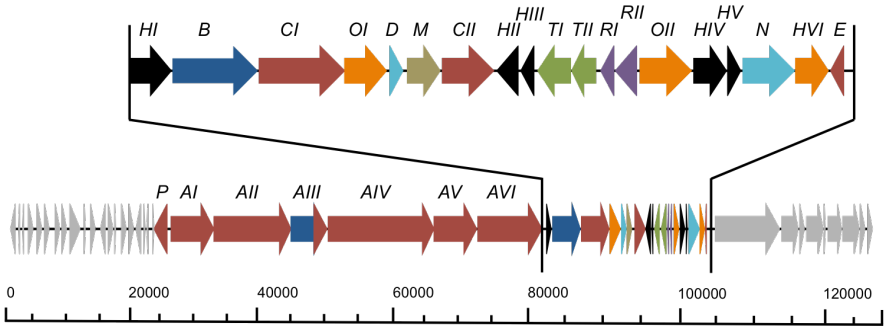


Figure 7 Schematic representation of the biosynthetic gene cluster of kirromycin. Red, PKS-related genes; Blue, NRPS-related genes; Black, hypothetical proteins; Orange, dehydrogenases and hydroxylases; Light blue, genes

involved in precursor supply; Light brown, O-methyltransferase; Light green, transport-related genes; Purple, resistance-related genes; Light grey, genes not involved in kirromycin biosynthesis. Modified from Weber *et al.* 2008[25].

Kirromycin is synthesised and assembled by a hybrid type I PKS/NRPS assembly line in which *kirAI, kirAII,* and *kirAIV – AVI* encode PKS modules, *kirB* an NRPS module, and finally *kirAIII* is composed of both NRPS- and PKS-related modules (**Fig. 8**). Based on the studies of Weber *et al.*[25], loading of the acetyl-CoA starter unit onto KirAI might be facilitated by the putative Gcn5-related N-acetyltransferase (GNAT) KirE, although no experimental data is available. Further extension of the precursor molecule is achieved from the PKS modules KirAII and KirAIV – AVI, which amounts to the addition of 15 extender units consisting of 13 malonyl-CoA, one methylmalonyl-CoA, and one ethylmalonyl-CoA extender unit. The two NRPS modules, KirAIII and KirB, account for incorporation of the glycine and β-alanine residue, respectively. Post-translational activation of the ACPs and PCPs is catalysed by the putative Sfp-type phosphopantetheinyl transferase (PPTase) encoded by *kirP*[51]. Besides its ability to activate both ACP and PCP, KirP also has a relaxed substrate specificity with respect to CoA substrates and has been found to accept both malonyl- and methylmalonyl-CoA.

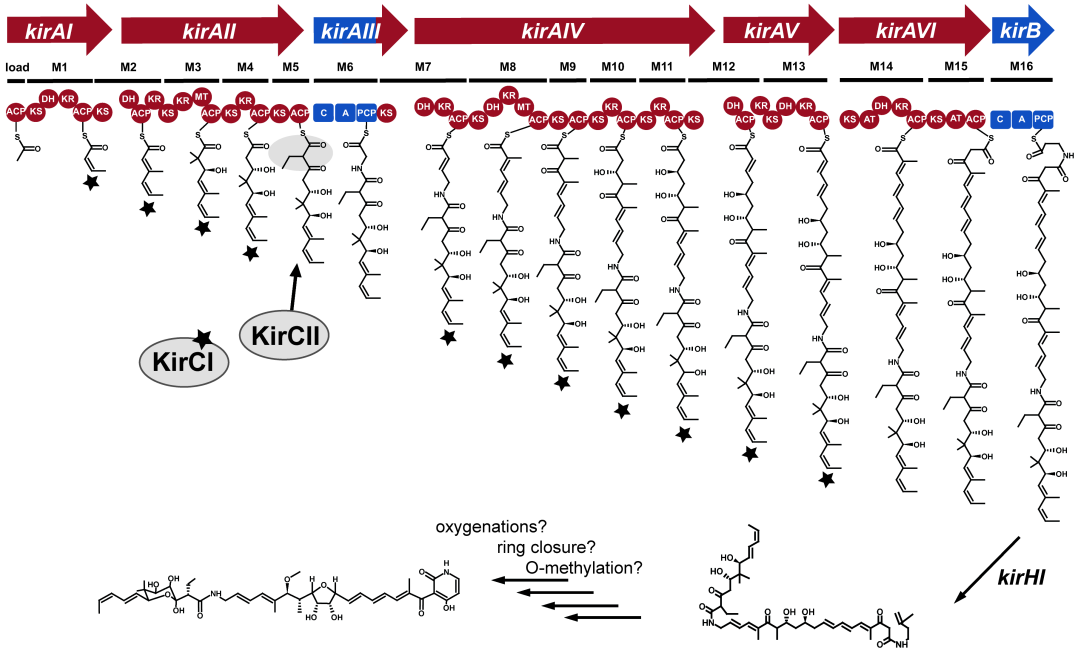


Figure 8 Schematic overview of the hybrid PKS type I/NRPS assembly line for kirromycin. ACP, acyl carrier protein; AT, acyltransferase; KS, ketosynthase; DH, dehydrogenase; KR, ketoreductase; ER, enoylreductase; MT, methyltransferase; C, condensation domain; A, adenylation domain; PCP, peptidyl carrier protein; KirCI/KirCII, *trans*-ATs; *kirHI*, Dieckmann cyclase; *kirM*, O-methyltransferase; kirHVI, phytanoyl-CoA dioxygenase; *kirOI*/*kirOII*, cytochrome P450-dependent hydroxylases. Modified from Weber *et al.* 2008[25]

Several interesting features of the kirromycin BGC have led to additional studies to elucidate the logic behind its biosynthesis. Most intriguingly, only KirAVI contains the classical *cis*-AT domains responsible for selecting and loading of extender units onto the ACPs. Instead, KirAI to KirAV are *trans*-AT PKSs, making kirromycin the first described biosynthetic pathway combining both types of PKS. The extender units for the *trans*-AT type PKSs are provided by the two discrete ATs, KirCI and KirCII. More specifically, KirCI accepts and loads malonyl-CoA extender units onto all modules except for M1 (starter unit) and M5, while KirCII is responsible for loading of ethylmalonyl-CoA onto ACP5 in M5. *In vitro* assays of both enzymes have enabled the characterisation of the enzyme specificities. Despite the three-domain architecture of KirCI (AT_1-AT_2-ER), only the AT_2 domain is responsible for loading of malonyl-CoA extender units onto ACPs in the kirromycin assembly line[27]. Unlike KirCI, which has a high specificity towards malonyl-CoA extender units, KirCII is more promiscuous when it comes to substrate loading, in that it was found to accept non-malonyl-CoA extender units, such as allyl- and propargyl-malonyl-CoA at the ACP5[26,52]. Here, the chemical nature of propargyl-kirromycin has allowed for its use in Copper(I)-dependent azide-alkyne cycloaddition (CuAAC), in popular term "click" chemistry, which allows for the exchange of functional groups on a molecule under mild aqueous conditions. In an eGFP transcription/translation *in vitro* assay of kirromycin, allyl-kirromycin, and coumarin-kirromycin, the latter generated by "click" chemistry, a lowered inhibition of EF-Tu was observed for coumarin-kirromycin (IC_{50} = 7.3 μM) compared to kirromycin (IC_{50} = 0.9 μM) (**Fig. 9**). This observation strengthened the incentive to use this polyketide derivatization tool for changing the pharmacokinetic properties of kirromycin[52].

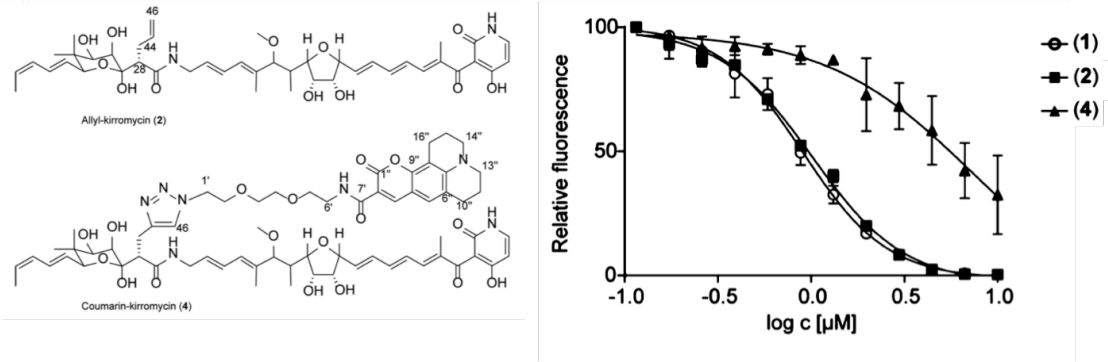


Figure 9 Structures of kirromycin analogues allyl-kirromycin (2) and coumarin-kirromycin (4) and the *in vitro* translation assay with eGFP inhibition as a result of increasing concentrations of kirromycin (1) and the two kirromycin analogues (2) and (3). Modified from Musiol-Kroll *et al.*[52]

Another interesting feature of the kirromycin assembly line is its missing TE domain. Instead, the precursor molecule of kirromycin is released from the megaenzymatic complex by KirHI, which was recently found to belong to a new class of Dieckmann cyclases responsible for tetramic acid and pyridine scaffold biosynthesis[53].

Although much is known about the core genes involved in the biosynthesis of kirromycin, knowledge about host resistance and tailoring steps remain scarce. From the initial studies of Weber *et al*.[25], no genes within the cluster boundaries could be assigned to be involved in resistance. Interestingly, *S. collinus* Tü 365 has a kirromycin-sensitive EF Tu, but as the strain does not suffer during production of the antibiotic, alternative resistance mechanisms must be present and are yet to be characterised[54]. In the case of the tailoring reactions, enzymes responsible for oxygenations, hydroxylations, formation of the tetrahydrofuran ring, and *O*-methylation are expected to be encoded by genes within the BGC of kirromycin as well, although they remain to be characterised.

Manuscripts

Manuscript I

Metabolic engineering of streptomycetes

Streptomycetes possess a huge potential to produce valuable natural products, however, many of the BGCs encoding the biosynthesis remain silent under standard growth conditions. Through the years, several approaches have been employed for unraveling the hidden potential of these gifted microorganisms. These include systematic manipulation of the fermentation conditions, also known as the One Strain-Many Compounds (OSMAC) approach[55], and mutagenesis procedures[56] to isolate novel mutants with improved production, increased tolerance to the end product, or with other traits rendering the strain superior to the parental strain.

In recent years, more systematic and rational approaches are being pursued in order to unlock the hidden potential of streptomycetes. One such focused approach presents itself with metabolic engineering, which builds on *de novo* assembly or modifications of metabolic pathways within a strain to improve flux towards a desired end product[57].

Metabolic engineering of well-established hosts such as *Escherichia coli* and *Saccharomyces cerevisiae* have already resulted in numerous successful examples of titer improvements for biofuels, bulk chemicals, proteins, and natural products[58–62]. Some key examples include the establishment and improvement of production of the erythromycin precursor 6-deoxyerythronolide B (6dEB) in *E. coli*[59] and the antimalerial drug precursor artemisinic acid in *S. cerevisiae*[58].

Despite the obvious benefits of using these model organisms as production hosts, they suffer from inherent limitations when it comes to production of the often complex secondary metabolites produced by actinomycetes and fungi. These limitations include incompatibility between the expression systems of the host and the natural producer, including insufficient precursor supplies, and poor growth of the production host caused by sensitivity towards the end product.

Heterologous hosts of streptomycetes origin overcome some of these limitations, of which *Streptomyces coelicolor* M1154*, Streptomyces avermitilis*, and *Streptomyces albus* J1074 are some noteworthy mentions[63]. More on this topic and the recent

advances made in the transition to systems metabolic engineering of streptomycetes for improved production of secondary metabolites are presented in **manuscript I**, a review entitled "Towards systems metabolic engineering of streptomycetes for secondary metabolites production", which was recently published in Biotechnology Journal[64].

Manuscript I

Towards systems metabolic engineering of streptomycetes for secondary metabolites production

Helene Lunde Robertsen[1], Tilmann Weber[1], Hyun Uk Kim[1,2], and Sang Yup Lee[1,2]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark.

[2]Department of Chemical and Biomolecular Engineering and BioInformatics Research Center, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea.

(Accepted review in Biotechnology Journal, published online November 13th 2017)

REVIEW

Systems Metabolic Engineering

Biotechnology Journal

www.biotechnology-journal.com

Toward Systems Metabolic Engineering of Streptomycetes for Secondary Metabolites Production

Helene Lunde Robertsen, Tilmann Weber, Hyun Uk Kim, and Sang Yup Lee**

Streptomycetes are known for their inherent ability to produce pharmaceutically relevant secondary metabolites. Discovery of medically useful, yet novel compounds has become a great challenge due to frequent rediscovery of known compounds and a consequent decline in the number of relevant clinical trials in the last decades. A paradigm shift took place when the first whole genome sequences of streptomycetes became available, from which silent or "cryptic" biosynthetic gene clusters (BGCs) were discovered. Cryptic BGCs reveal a so far untapped potential of the microorganisms for the production of novel compounds, which has spurred new efforts in understanding the complex regulation between primary and secondary metabolism. This new trend has been accompanied with development of new computational resources (genome and compound mining tools), generation of various high-quality omics data, establishment of molecular tools, and other strain engineering strategies. They all come together to enable systems metabolic engineering of streptomycetes, allowing more systematic and efficient strain development. In this review, the authors present recent progresses within systems metabolic engineering of streptomycetes for uncovering their hidden potential to produce novel compounds and for the improved production of secondary metabolites.

1. Introduction

Streptomycetes are filamentous gram-positive bacteria predominantly found in soil and water environments. The bacteria are recognized for their capabilities to produce secondary metabolites with clinically relevant applications, some of which include antibiotics such as streptomycin and daptomycin, anthelmintic compound avermectin, immunosuppressant tacrolimus (FK-506), and anti-cancer agents bleomycin and doxorubicin (for review, see Ref. [1]). Despite the proven clinical effects of secondary metabolites, drug discovery in streptomycetes has witnessed a decline in success rate ever since the golden age of antibiotics experienced from late 1940s to 1960s.[2] A paradigm shift begun in the early 2000s when the complete genome sequences of a model organism *Streptomyces coelicolor*[3] and an industrial strain *Streptomyces avermitilis*[4] were published. In the case of *S. coelicolor*, the 8.6 Mb linear chromosome was found to harbor several cryptic or silent biosynthetic gene clusters (BGCs) in addition to already known BGCs of actinorhodin, undecylprodigiosin, and calcium-dependent antibiotic (CDA). These silent BGCs showed indications of possible biosynthesis of secondary metabolites, but interestingly, no corresponding compounds were detected under standard laboratory growth conditions. In the following years, many of these silent BGCs were successfully activated and their corresponding secondary metabolites were characterized, including iron-chelators desferrioxamine[5] and coelichelin,[6,7] sesquiterpene antibiotic albaflavenone,[8] and polyketide (PK) alkaloid coelimycin.[9] The initial successes in identifying novel secondary metabolites from silent BGCs have led to the subsequent increases in the genome sequencing of streptomycetes in the last decade (**Figure 1**).

Availability of genome data has spurred advances in relevant technologies covering computational resources (in particular, genome and compound mining tools), high-throughput (omics) techniques as well as molecular tools. Genome mining tools such as antibiotics and Secondary Metabolites Analysis SHell (antiSMASH)[10–12] and Prediction Informatics for Secondary Metabolomes (PRISM)[13,14] have greatly improved our ability to survey the genomic potential of strains, and allowed for initial prioritization of engineering efforts. Also, omics techniques such as transcriptomics and proteomics have contributed to understanding the complex regulatory networks employed by streptomycetes to balance primary and secondary metabolism.[15,16] Availability of such genome and expression data has further enabled construction of genome-scale metabolic models (GSMMs) of streptomycetes to describe metabolic pathways and predict optimal routes for the production of secondary metabolites.[17]

With high-quality whole genome sequences and an array of suitable genome/compound mining and molecular cloning tools readily at hand, the secondary metabolite research is now

H. L. Robertsen, Dr. T. Weber, Prof. S. Y. Lee
The Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark
2800 Kongens Lyngby, Denmark
E-mail: tiwe@biosustain.dtu.dk; leesy@kaist.ac.kr

Dr. H. U. Kim, Prof. S. Y. Lee
Department of Chemical and Biomolecular Engineering
(BK21 Plus Program)
Korea Advanced Institute of Science and Technology (KAIST),
Yuseong-gu, Daejeon 306-701, Republic of Korea

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/biot.201700465.

DOI: 10.1002/biot.201700465

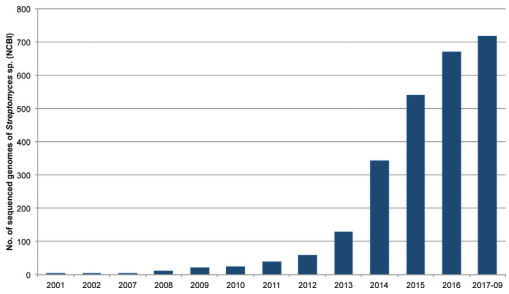


Figure 1. Cumulative number of whole genome sequences of *Streptomyces* species published in National Center for Biotechnology Information (NCBI) from 2001 to September 2017.

moving toward systems metabolic engineering of streptomycetes in order to fully harness their potential to produce medically valuable secondary metabolites.[18] Systems metabolic engineering brings in recent developments experienced in the fields of systems and synthetic biology, and provides a more systematic and efficient approaches for strain engineering.[19] However, streptomycetes and other secondary metabolite-producing microbes require additional sophisticated analyses concerned with secondary metabolite BGCs before strains are engineered at the systems level for the novel compound production and the improved production titers.[17] To this end, here we review recent progress within the field of systems metabolic engineering of streptomycetes with focuses on the additional sophisticated analyses unique to streptomycetes as well as recent successful studies on streptomycetes engineering (**Figure 2**).

2. Unique Considerations for Systems Metabolic Engineering of Streptomycetes

The established framework of systems metabolic engineering[19] can serve as a guideline for the optimal production of secondary metabolites using streptomycetes (**Figure 3**). However, systems metabolic engineering of streptomycetes has additional special considerations that are not necessarily relevant to popular model organisms such as *Escherichia coli* and *Saccharomyces cerevisiae*. In this context, we first focus on these "unique considerations" for streptomycetes, including characterization of BGCs, their encoding secondary metabolites and regulations involved in the secondary metabolite biosynthesis as well as molecular tools specifically adapted for the streptomycetes engineering.

2.1. Characterization of BGCs and Their Secondary Metabolites Through Computational Resources

Due to the complex nature of secondary metabolism, computational resources are needed to better understand working mechanisms of secondary metabolisms many of which are still unknown. Secondary metabolite biosynthetic pathways are also not sufficiently covered by existing general metabolic databases

Helene Lunde Robertsen received her M.Sc. degree in Biotechnology at the Technical University of Denmark in 2014. During her master thesis, she worked on metabolic engineering of *Saccharomyces cerevisiae* for the optimization of capsaicin production. She is currently pursuing a Ph.D. in Bioengineering and Synthetic Biology at the Novo Nordisk Foundation Center for Biosustainability. Her research involves genetic engineering of actinomycetes for discovery and production of novel antibiotics.

Tilmann Weber is Co-Principal Investigator of the New Bioactive Compound section at the Novo Nordisk Foundation Center for Biosustainability of the Technical University of Denmark. He is interested in integrating bioinformatics, genome mining, and systems biology approaches into Natural Products discovery and characterization and thus bridging the in silico and in vivo world. He obtained his Ph.D. (with Prof. Dr. Wolfgang Wohlleben) and his habilitation at the Eberhard Karls University Tübingen, Germany. Tilmann is Associate Editor of Synthetic and Systems Biotechnology and Editorial Board member of Scientific Reports and Cell Chemical Biology.

Hyun Uk Kim is Research Fellow at Korea Advanced Institute of Science and Technology (KAIST), South Korea. His research field lies in systems biology, biochemical and metabolic engineering, and drug targeting and discovery. One of his current studies is focused on development of systems approaches to produce and characterize natural products. He earned his Ph.D. at KAIST under the supervision of Prof. Sang Yup Lee.

Sang Yup Lee is Distinguished Professor and Dean of KAIST Institutes at KAIST. He is also Scientific Director at the Novo Nordisk Foundation Center for Biosustainability, DTU. He has published more than 560 journal papers, 74 books/book chapters, and more than 630 patents, largely in the field of metabolic engineering and industrial biotechnology. He has been elected to several academies worldwide including National Academy of Engineering USA and National Academy of Sciences USA. He has received numerous awards.

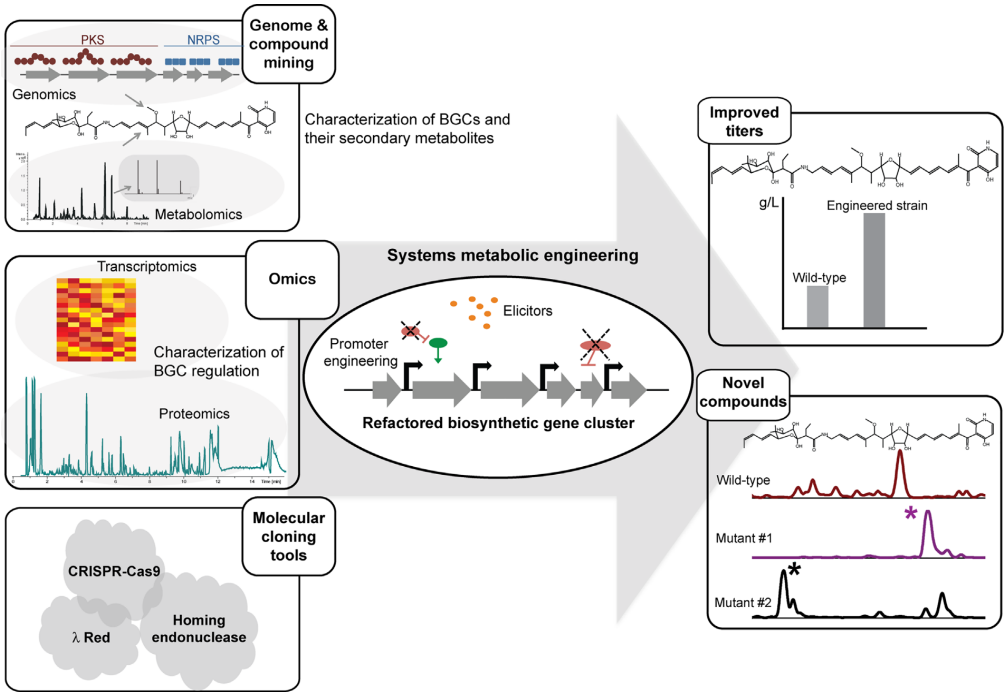


Figure 2. Schematic representation of the workflow of systems metabolic engineering for streptomycetes. Characterization of BGCs, their encoding secondary metabolites and biosynthesis regulations involved is critical in engineering streptomycetes for the production of novel compounds and the higher yields of a known end product. Genome and compound mining, omics techniques, molecular tools as well as the use of elicitors play important roles in this process.

(e.g., KEGG and MetaCyc). Fortunately, advances in high-quality genome sequences of *Streptomyces* isolates (Figure 1) and other secondary metabolite-producing organisms have sparked the development of such specialized computational resources, including databases and genome/compound mining tools, and the representative ones are presented herein. First, BGC databases allow for easy access and evaluation of curated and predicted BGCs, such as the Minimum Information on Biosynthetic Gene Clusters (MIBiG),[20] the Integrated Microbial Genomes Atlas of Biosynthetic gene Clusters (IMG-ABC),[21] and the antiSMASH database.[22] For the secondary metabolite compounds, several comprehensive databases exist as well, including NORINE[23] specifically for non-ribosomal peptides (NRPs), Antibiotic'ome[24] for the predicted molecular targets of antibiotics, and StreptomeDB,[25] which presents information on compounds produced by streptomycetes.

Second, genome mining tools continue to develop with more features, which allow for predicting and assigning functions to enzymes involved in the biosynthesis of secondary metabolites. antiSMASH[12] and PRISM[14] are two of the most well-established tools, which both base their predictions on identification of signature genes or domains known to be specific for secondary metabolite BGCs.[20] Another recently released approach called EvoMining employs phylogenomic analysis to identify repurposed enzymes, originating from primary metabolism, which has been recruited to secondary metabolism for the secondary metabolite biosynthesis.[26] Using this EvoMining approach, Cruz-Morales et al. identified a new clade of biosynthetic enzymes involved in the biosynthesis of arseno-organic metabolites.[26] In addition, an alternative way of mining BGCs is found in the Antibiotics Resistant Target Seeker (ARTS) (https://arts2.ziemertlab.com), which evaluates BGCs based on self-resistance mechanisms of antibiotic producers.[27] ARTS scans genomes for promising BGCs based on known resistances and the presence of duplicated, co-localized housekeeping genes that display an evidence of horizontal gene transfer. In the context of systems metabolic engineering, the identified resistance genes can be used as probes to find novel BGCs without requiring detailed information on the BGC. The Gene Cluster Family (GCF) network (http://www.igb.illinois.edu/labs/metcalf/gcf/index.html)[28] is another useful approach for BGC identification, which groups gene clusters and assigns functions of previously uncharacterized BGCs.

For the compound mining, further prioritization of BGCs in promising strains can be achieved through cheminformatic analyses of the metabolome data generated from liquid chromatography (LC) tandem mass spectrometry (MS/MS) or

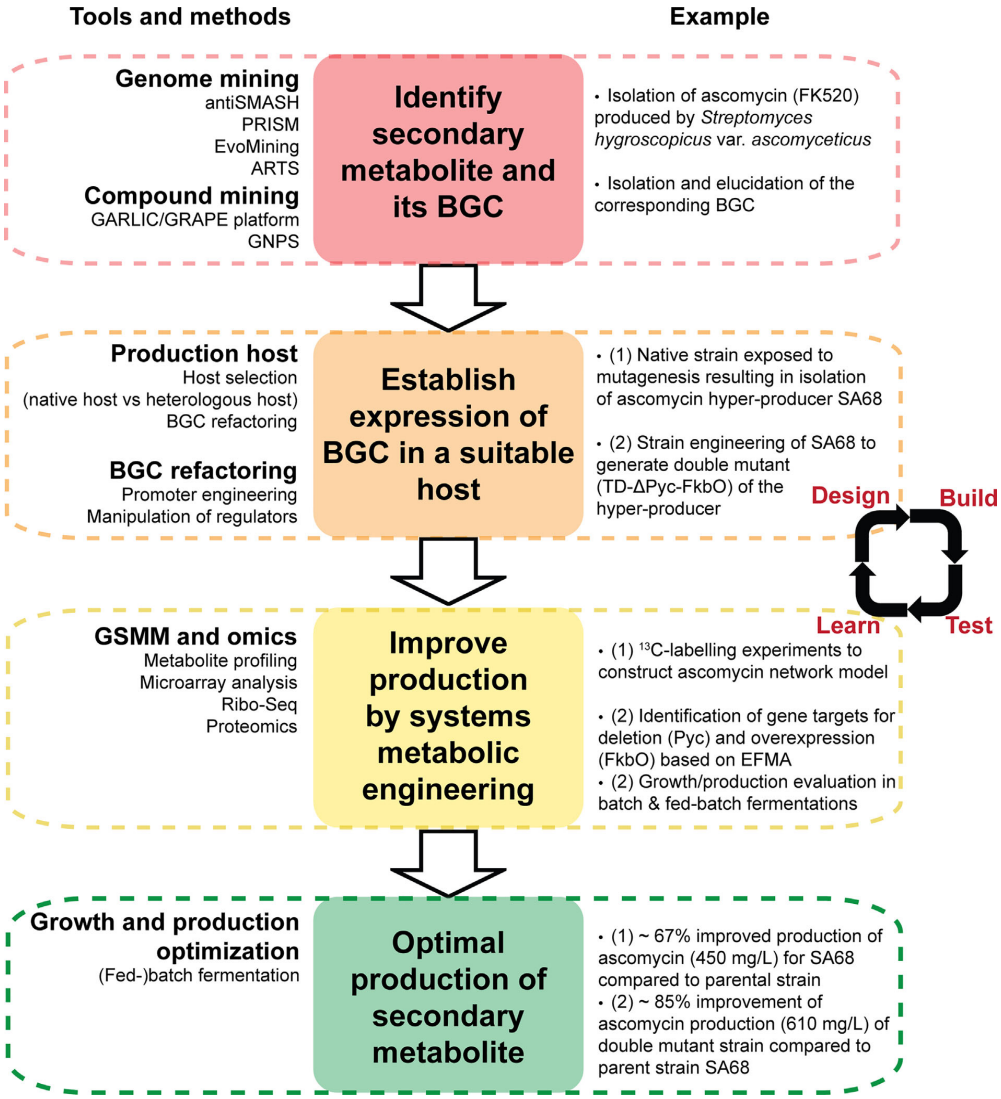


Figure 3. Workflow of systems metabolic engineering for the optimal production of secondary metabolites. The workflow shown includes tools and methods used within the individual steps. The workflow can be applied in an iterative fashion as illustrated by the design-build-test-learn cycle. The use of the entire workflow is exemplified by the production of ascomycin using *Streptomyces hygroscopicus* var. *ascomyceticus*. Here, (1) and (2) illustrates the two separate rounds of engineering applied for the optimization of ascomycin production.

nuclear magnetic resonance (NMR).[13,29] Upon LC-MS/MS-guided identification of new compounds, following compound mining (or cheminformatic) tools enable automated compound characterization and dereplication: Global Alignment for natuRaL-products chemInformatiCs (GARLIC) in combination with Generalized Retrobiosynthetic Assembly Prediction Engine (GRAPE) platform[30]; and the open-access MS/MS database Global Natural Products Social Molecular Networking (GNPS).[31] The GARLIC pipeline allows for both linking orphaned products to their corresponding BGCs and for identification of novel compounds by aligning bacterial PK and NRP BGCs predicted by PRISM to a comprehensive library of natural products from GRAPE.[30] The applicability of this pipeline consisting of PRISM, GRAPE, and GARLIC was

demonstrated by identifying a new compound, potensibactin, isolated from *Nocardiopsis potens* DSM 45234, and also by linking three previously orphaned natural products, lucensomycin, octacosaminin, and bogorol, to their respective BGCs in *Streptomyces achromogenes* NRRL 3125, *Amycolatopsis* sp. NAM 50, and *Brevibacillus laterosporus* DSM 25, respectively. Meanwhile, the GNPS platform contains a comprehensive compound data collection curated by the natural product community.[31] As this community-driven knowledge-sharing database constantly expands, it allows for continuous screening of newly uploaded data sets against the established database. Besides its applicability in the discovery and characterization of natural products and their BGCs, the network can be used to identify structural analogs of a known molecule as exemplified from the analysis of a broad-spectrum antibiotic stenothricin from *Streptomyces roseosporus*. Based on the molecular networking and dereplication features of the platform, GNPS identified a subnetwork of five stenothricin analogs produced by *Streptomyces* sp. DSM5940, all with different structural properties compared to the original molecule stenothricin.

More detailed descriptions beyond these computational resources discussed herein can be found from several review papers.[32–34] In addition, to aid scientists within the field in prioritizing the individual tools, The Secondary Metabolite Bioinformatics Portal (SMBP) at http://www.secondarymetabolites.org was recently launched.[35] For each tool or database, SMBP provides the users with links and short descriptions, allowing for quick and easy browsing. SMBP can especially be useful for metabolic engineers who plan to produce secondary metabolites, but are not familiar with secondary metabolism biochemistry.

2.2. Characterization of the Regulation of Secondary Metabolites Biosynthesis in Streptomycetes – Use of Omics Techniques as an Example

The data obtained from genome mining often provides little information on the regulation of secondary metabolites biosynthesis in streptomycetes. An additional level of information from omics techniques can be used to clarify the complex regulations involved.[36] Transcriptomics is the most accessible and most frequently adopted omics approach which monitors changes in the gene expression levels over a course of time, or as a result of external stimuli or genetic manipulation.[16] In this context, Nieselt et al.[37] conducted microarray analyses of time-point samples of *S. coelicolor* M145 taken during a 60 h fermentation, and established changes in the expression profiles of clustered genes during the metabolic switch from primary to secondary metabolism. Here, the general change in expression profiles observed between the 35 and 36 h especially marked the metabolic switch, which was an interval for the depletion of phosphate from the medium. More recently, multi-omics techniques were deployed to determine the transcriptional and translational landscape of *S. coelicolor* M145 using differential RNA-sequencing (dRNA-seq) and strand-specific RNA-seq (ssRNA-seq) first, followed by ribosome profiling (Ribo-seq).[38] In particular, the use of Ribo-seq shed light on the level of translational control on genes involved in both primary and secondary metabolism, and revealed a general decline in the translation efficiency of BGCs in *S. coelicolor* after the cell's transition to growth phase. Furthermore, translation efficiency was negatively correlated with transcription, hence revealing the extra level of regulation to account for when embarking on the quest of strain optimization.

In addition to transcriptomics, proteomics can provide additional and complementary information on regulations associated with the transition to antibiotic production. Proteome data of *S. coelicolor* M145 were collected at three developmental stages ranging from compartmentalized mycelium (12 h) to the multinucleated mycelium (24 h and 72 h), which revealed that hypha differentiation from substrate to aerial hypha was found to be correlated with a change in proteome composition.[39] Based on LC-MS/MS, 626 proteins were identified, of which 345 proteins were further quantified to reveal that proteins involved in primary metabolism were the most abundant in the cells in the compartmentalized mycelium, whereas proteins involved in secondary metabolism were found to predominate the cells in the multinucleated mycelium. Such omics studies will allow us to better understand metabolic characteristics and complex regulations involved in biosynthesis of desired secondary metabolites.

2.3. Molecular Tools for Streptomycetes Engineering

A remarkable progress recently made in the computational field has been accompanied with concurrent advances in molecular tools in order to overcome challenges associated with streptomycetes engineering: high GC contents, relatively long cultivation times and occasional (or frequent in some strains) reluctance towards introduction of recombinant DNA.[40] To circumvent the limitations encountered when using the standard restriction- and ligation-based cloning for genetic manipulation, the *E. coli*-derived λ Red system was applied for faster and more efficient gene disruption, in-frame deletion, gene replacement, and refactoring of entire gene clusters in *Streptomyces* strains.[41] To aid genetic engineering effort even further in streptomycetes, an alternative to this classical method was reported using the *S. cerevisiae* I-SceI homing endonuclease for DNA double strand break (DSB)-based genome editing.[42,43] This meganuclease differs from conventional restriction enzymes by its longer recognition site of 18 bp and with a lower risk of off-targets for gene editing purposes. Using a two-plasmid system, the I-SceI recognition sequence and a suitable resistance marker are introduced into a defined site on the genome of the recipient streptomycete via a single crossover event using the first plasmid. Double crossover is induced upon introduction of the second plasmid, which harbors the synthetic codon-optimized I-SceI under the control of either the constitutive *ermE** or thiostrepton-inducible *tipA* promoter. To facilitate marker-free gene deletions, plasmids derived from pSG5[44] with temperature sensitive replicons can be used. When using the I-SceI system for gene replacement in *Streptomyces* sp. Tü 6071, I-SceI-mediated homologous recombination (HR) was found to be 25 times more efficient than spontaneously occurring HR, hence, establishing its use for more systematic gene editing in streptomycetes.[42]

More recently, an even more efficient and flexible system for genome engineering in streptomycetes has made its way to the field, namely the type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated protein (Cas9) system from *Streptococcus pyogenes*. The CRISPR-Cas9 technique can be applied for deletion of genes and gene clusters[45–47] in a multiplex manner,[45] reversible gene expression control,[47] and induction.[48] The optimized system relies on two components: the single guide RNA (sgRNA), which is a synthetic RNA consisting of a CRISPR RNA (crRNA) and *trans*-activating crRNA (tracrRNA) complex, and is required for guiding Cas9 to modify the targeted genome sequence; and the endonuclease Cas9 or catalytically inactive dCas9, which upon interaction with the sgRNA, scans the genome for protospacer adjacent motif (PAM) sequences and the corresponding 20 bp recognition site. Upon recognition of the target DNA sequence, Cas9 binds to and cleaves the specific site in the genome whereas the inactive dCas9 will bind to the target DNA without sequence modifications. Here, the dCas9 presents an easy-to-clone tool for reversible control of gene expression, and thus is useful for investigating regulation without the risk of off-target effects or subjecting the cells to the stress associated with gene replacements or mutagenesis procedures.[47] CRISPR-Cas9-mediated mutagenesis can be used in the absence of templates for HR. In this case, the repair mechanism is dependent on the non-homologous end joining (NHEJ) of the DNA in *Streptomyces* species. However, efficiency and specificity are greatly improved if a homology template for HR is provided.[45–47] This has been demonstrated in independent studies, in which Cobb et al.[45] and Huang et al.[46] reported 60≈100% and 70≈100% efficiencies, respectively, whereas Tong et al.[47] reported near 100% efficiency for their studies on the actinorhodin gene cluster in *S. coelicolor*.

For specific applications, the CRISPR-Cas9 system has proven suitable for identification and examination of putative genes and BGCs for secondary metabolism in several *Streptomyces* species. One example is the use of the system for discovery and examination of the BGC encoding the antibiotics formicamycins in a new isolate *Streptomyces formicae*.[49] Using the pCRISPomyces-2 vector, which harbors both the codon-optimized Cas9 from *S. pyogenes* and the sgRNA cassette, the only type II polyketide synthase (PKS) BGC in *S. formicae* was deleted, resulting in a mutant incapable of producing any formicamycins. Using the same vector, the authors proceeded to elucidate the biosynthetic pathway by knocking out single genes in the putative BGC, and ultimately proposed a preliminary biosynthetic route for the production of this group of antibiotics.[49] Similarly, CRISPR-Cas9 was used to determine that the LuxR family cluster-situated regulator FscRI was necessary for the activation of antimycins biosynthesis in *Streptomyces albus* S4.[50] Using this information, heterologous production of antimycins in *S. coelicolor* M1146 was achieved when the BGC was co-expressed with its activator FscRI. The general versatility of the CRISPR-Cas9 system has been further exemplified by its application for genome editing of rare actinomycetes, one of which is *Actinoplanes* sp. SE50/110.[51]

The CRISPR-Cas9 was applied in an alternative fashion for the knock-in strategy in order to activate silent BGCs in multiple *Streptomyces* strains.[48] Replacing native promoters with the strong, constitutive promoter *kasOp** using the CRISPR-Cas9 system led to the production of new compounds in *S. roseosporus*, *Streptomyces venezuelae*, and *Streptomyces viridochromogenes*. Furthermore, the compound isolated from *S. viridochromogenes* was found to be a novel pigmented PK produced from an otherwise silent BGC of the type II PKS. Although this activation strategy bears striking similarity to a technique previously reported for the activation of five gene clusters in *S. albus* J1074,[52] the incentive to use the CRISPR-Cas9 system is the increased efficiency of HRs, presenting an improved method for genetic manipulation of strains suffering from the low efficiency of natural recombination.

The use of synthetic or natural strong promoters has also been proven useful in activating BGCs with low or no expression levels.[53–55] Based on the *xylE* reporter gene and RNA-seq, 32 promoters from *S. albus* J1034 were identified and characterized, revealing 10 promoters with strengths 200≈1300% higher than that reported for the standard, constitutive promoter *ermE** p.[55] Using a similar screening approach, Siegl et al.[54] evaluated their collection of synthetic promoters, and reported strengths of 2≈319% in comparison with that of *ermE** p.

In addition to the newly adapted molecular tools discussed above, various systems metabolic engineering strategies for the model organisms can also be deployed to refactor secondary metabolites biosynthesis in native hosts, including removal of competing pathways to change the flux distribution or increase the pool of precursors, deletion of repressors, and overexpression of activators.[56] With further advances in the field, systems metabolic engineering strategies will also be employed for establishment and optimization of secondary metabolites production in heterologous hosts in the future (see Section 2.4).

2.4. Additional approaches that facilitate streptomycetes engineering

Besides the molecular tools discussed above, several additional approaches exist, which may further facilitate streptomycetes enginering and the secondary metabolite production. First, if the native host is reluctant toward any genome engineering efforts, heterologous expression might prove a helpful alternative.[40,57] Strains such as *S. albus*, *S. avermitilis*, *S. coelicolor*, and *Streptomyces lividans* are readily amenable to cloning and expression, and therefore, their use as heterologous hosts circumvents native host limitations.[58] Also, the construction of genome-minimized hosts, which have had one or more BGCs removed from their genomes, might improve titers of a given product by avoiding the use of common precursors for other secondary metabolites.[56,58] One such example is the minimized host *S. coelicolor* M1154 which had four of its BGCs deleted in addition to the introduction of point mutations in *rpoB* and *rpsL*, and is known to have an improved antibiotic production performance. Using this host, Gomez-Escribano and Bibb reported 30- and 40-fold increases in the production titers of congocidine and chloramphenicol, respectively, compared to the control strain *S. coelicolor* M145 as the production host.[59] An even further reduction of the *S. coelicolor* genome was obtained by Zhou et al.[60] who, using PCR-targeting of cosmids for the gene disruption, sequentially deleted all the 10 BGCs encoding

PKS and non-ribosomal peptide synthetases (NRPS) in addition to a 900-kb subtelomeric region. A series of mutants were generated as a result, one of which had 14% of its genome deleted. Furthermore, a physiological comparison of the constructed genome-minimized mutants to the wild-type showed no differences in growth rates. Hence, this strategy can provide new hosts suitable for heterologous expression of a wide variety of BGCs.

The development of a small molecule activation approach could also prove helpful when working with the troublesome isolates. In this regard, chemical elicitation which uses small molecules to induce the metabolites production in a given strain has been used with positive outcomes.[61,62] Using the GFP- and LacZ-based reporter systems to monitor activation of gene expression, a library of 640 elicitors was screened to find nine of them eliciting production of two otherwise silent BGCs in *Burkholderia thailandensis* E264.[61] Similarly, in a study on the effects of rare earth elements on secondary metabolite production in *S. coelicolor,* scandium and lanthanum were found to induce gene expression for the actinorhodin BGC in addition to four otherwise silent BGCs.[63] More recently, a library of 30 569 small molecules was screened for their effects on actinorhodin production in *S. coelicolor* M145.[64] This led to the isolation of a group of four antibiotic-remodeling compounds (ARCs), of which ARC2 was found capable of eliciting secondary metabolism in related *Streptomyces* species as well. Following these findings, the synthetic derivative Cl-ARC was used for the stronger elicitation of BGCs encoding molecules at a low abundance in a panel of fifty *Actinomycete* strains.[65] This resulted in the identification of several CI-ARC-induced compounds including oxohygrolidin from *Streptomyces ghanaensis,* not previously known to produce this molecule.[66] Furthermore, based on bioactivity-guided assays, it was shown that while the extracts of the elicitor-treated *S. ghanaensis* failed to provide detectable inhibition, purified oxohydrolidin displayed activity against *S. cerevisiae* Y7092. The latter finding provides strong incentives to focus future work on purified compounds in order to avoid masking of low abundance molecules in complex sample mixtures. Furthermore, such systematic analyses show a great promise in high-throughput screening approaches in that they allow for both screening multiple elicitors on one strain, or, as reported by Craney et al.,[64] screening one elicitor molecule against a collection of strains. The major limitation to such an approach, however, remains in the massive amount of samples generated, which all require time-consuming subsequent analytics studies.

3. Recent Examples of Systems Metabolic Engineering of Streptomycetes for the Optimized Production of Secondary Metabolites

The framework of systems metabolic engineering continues to be adopted to improve the production of secondary metabolites in streptomycetes, as seen from the many recent successful examples reported (**Table 1**). Some noteworthy examples include the production of tacrolimus (FK506) using *Streptomyces tsukubaensis*,[67–69] ascomycin (FK520) using *Streptomyces hygroscopicus* var. *ascomyceticus*,[70–72] avermectin using *S. avermitilis*,[73] and pristinamycin I (PI) and II (PII) using *Streptomyces pristinaespiralis*.[74,75]

To solve the issue of low production yield of tacrolimus in the natural producer, Wang et al.[67] first determined the intracellular response of *S. tsukubaensis* to exogenous feeding of four precursors known to promote the tacrolimus biosynthesis. Using a weighted correlation network analysis (WGCNA) on a dataset containing 93 different intracellular metabolites measured with gas chromatography (GC)-MS and LC-MS/MS, metabolites highly associated with the tacrolimus biosynthesis were identified. Furthermore, time-point sampling led to the identification of three pronounced pathways involved in the tacrolimus biosynthesis, namely pentose phosphate, shikimate, and aspartate pathways. Using this information and the GSMM of *S. tsukubaensis,* the genes *aroC* and *dapA*, involved in the biosynthesis of chorismate and lysine, respectively, were identified for overexpression, and the effects of overexpression of these genes were experimentally validated. As a result, a mutant overexpressing the *aroC* and *dapA* genes produced 1.64-fold higher yields of tacrolimus, compared to the wild-type strain.

Ascomycin, an ethyl analog of tacrolimus, is another important immunosuppressant. To optimize the production of this compound, the natural producer *S. hygroscopicus* var. *ascomyceticus* has undergone several rounds of metabolic engineering. Early strain development efforts have so far included femtosecond laser irradiation mutagenesis combined with a shikimic acid enduring screening. Although shikimic acid was found to have a positive effect on the ascomycin production, the mutant strain FS35, from the mutagenesis screen, also displayed high sensitivity toward this substrate. To overcome this negative effect, an endurance screening with shikimic acid was carried out, resulting in the isolation of a shikimic acid-resistant mutant strain SA68 with ascomycin yield of 330 mg L^{-1} (270 mg L^{-1} for the parent strain).[70,71] More recently, systems metabolic engineering[72] was employed to obtain a high-yield ascomycin producing strain, which showed 84.8% titer improvement, compared to the parent strain. Prior to the strain engineering, ^{13}C-labelling experiments and metabolic flux analysis (MFA) were employed in parallel to construct and validate an ascomycin metabolic network. Elementary flux mode analysis (EFMA) on the model allowed the identification of chorismatase FkbO and pyruvate carboxylase Pyc as overexpression and inactivation targets, respectively. Furthermore, from the fed-batch fermentation of the high-producer strain TD-ΔPyc-FkbO, ethylmalonyl-CoA was found to be limited toward the end of fermentation, which opens up for future engineering efforts.

Similarly, industrial producer strains have undergone extensive systems metabolic engineering to improve their production titers of a group of avermectins.[73] As avermectin biosynthesis is under the tight control of several regulators, much focus has been put on relieving such regulators. The pathway-specific regulator AveR, which is involved in the activation of avermectin biosynthesis, is itself subject to both negative and positive regulation.[73] Here, the two TetR family transcriptional regulators (TFRs) encoded by the genes SAV576 and SAV577 were found to indirectly downregulate the avermectin biosynthesis through binding to the promoter region of SAV575; SAV575 is known to be involved in providing acetate and propionate extender units required for the avermectin production.[76,77] Using a combination of microarray analysis,

Table 1. Examples of metabolic engineering strategies used for the optimization of secondary metabolite production in streptomycete hosts.

Secondary metabolite	*Streptomyces* host	Metabolic engineering strategies	Highest yield increase (compared to control strain)	References
Actinorhodin	*Streptomyces coelicolor* A3(2)	Construction of two comprehensive GSMMs: iIB711 and iMK1208. The latest iMK1208 comprises 1208 genes, 1643 reactions, and 1246 metabolites	52-fold	[83–85]
		Construction of a mutant strain with improved yields		
		^{13}C-MFA and transcriptional analysis of mutant and parental strains to compare their fluxes and gene expression patterns, respectively		
		FSEOF on iMK1208 to identify additional gene target for overexpression and subsequent experimental verification		
Ascomycin	*Streptomyces hygroscopicus* var. *ascomyceticus*	Parallel ^{13}C labeling and MFA to construct metabolic network model	84.8%	[70–72]
		EFMA on the ascomycin network model for target predictions		
		Strain engineering by overexpression and inactivation of genes		
		Addition of resin HP20 in the growth medium		
Avermectin	*Streptomyces avermitilis*	Increasing flux through precursor pathways	3-fold	[73,76,77]
		Microarray analysis and ChIP assays for the analysis of expression levels		
		Identification and deletion of regulators with inhibitory effect on production		
Daptomycin	*Streptomyces roseosporus* LC-511	MFA to identify potential bottlenecks in biosynthesis	43.2%	[86,87]
		Identification of three genes for overexpression to increase fluxes through rate-limiting pathway		
		Construction of a triple overexpression mutant with improved titers		
		Transcriptional analysis to evaluate gene expression patterns of parental and mutant strains		
		Addition of glucose to fed-batch fermentations of the triple mutant strain to improve yields further		
Pristinamycin	*Streptomyces pristinaespiralis*	Transcriptional analysis by RT-PCR for the evaluation of regulatory network involved in biosynthesis	2.4-fold (PI)	[74,75]
		Strain engineering by introduction of extra copies of the BGC, deletion of repressors, and overexpression of activators	5.26-fold (PII)	
		Addition of resins to bioreactors		
Rapamycin	*Streptomyces hygroscopicus* ATCC 29253	Construction of a GSMM with 1003 reactions and 711 metabolites	142.3%	[88]
		FBA and MOMA analyses to evaluate gene targets for improving production		
		Construction of a mutant strain with one gene knock-out and overexpression of two genes showing higher fluxes through key primary metabolism pathways		
Spinosad	*Streptomyces albus* J1074	Construction of a *Saccharopolyspora spinosa* NRRL 18395 bacterial artificial chromosome library	1000-fold	[89]
		Establishment of heterologous expression of spinosyn BGC in *Streptomyces lividans* TK24 and *S. albus* J1074		
		Transcriptional and translational (proteomics) analyses to evaluate heterologous and native hosts		
		Metabolomics to identify the best suited candidate host for the expression of spinosyn BGC		
		Pathway refactoring including promoter engineering and overexpression of synthetic modules in *S. albus* mutant		

(*Continued*)

Table 1. (*Continued*)

Secondary metabolite	*Streptomyces* host	Metabolic engineering strategies	Highest yield increase (compared to control strain)	References
Tacrolimus	*Streptomyces tsukubaensis*	Proteomic and metabolomic analyses of an overproducer strain fed with soybean oil	1.64-fold	[67–69]
		WGCNA for the identification of pronounced precursor pathways		
		Identification of key limiting steps in the pronounced pathways using a GSMM		
		Gene overexpression to increase fluxes through two precursor pathways		

genetic studies, and chromatin immunoprecipitation (ChIP) assays, the expression levels of the TFRs and SAV575 were determined for both mutant and wild-type strains as well as the binding sites of SAV575 for the TFRs. This observation allowed for proposing the regulatory role of the TFRs in the avermectin production. Based on this information, a double deletion mutant (ΔSAV576 and ΔSAV577) was constructed, which showed ≈3-fold higher production of avermectin, compared to the wild-type.[77]

A more recent example of using a combinatorial metabolic engineering approach has been presented for the production of streptogramin-like antibiotic pristinamycin, which consists of two chemically unrelated compounds PI and PII.[78] As the native host *S. pristinaespiralis* suffers from low yields, additional copies of the two BGCs responsible for the production of PI and PII were introduced in addition to the deletion of repressors.[74,75]. The metabolic engineering efforts were strongly aided from previous findings of the regulatory cascade governing the biosynthesis of both PI and PII in *S. pristinaespiralis*.[79] For PII, the deletion of the two cluster-specific repressors PapR3 and PapR5 along with overexpression of the activators PapR4 and PapR6 resulted in a 1.5-fold higher production of PII in a mutant harboring an additional copy of the PII BGC, compared to the wild-type. Furthermore, the addition of resins to relieve both feedback inhibition and toxicity of PII resulted in the 5.26-fold higher production, compared to wild-type, when grown in 5 L bioreactors.[74] For PI, the highest production was observed in a mutant in which the BGC of PII was removed, an additional copy of the PI BGC introduced, and the repressor PapR3 deleted. Besides the 2.4-fold increased production of PI compared to the wild-type, interestingly, the mutant with only the PII BGC removed showed 20–40% lowered PI production, compared to the wild-type, revealing a possible role of PII as a coactivator or inducer of PI production.[75] Further characterization efforts are needed to unravel the regulatory mechanism behind this observation and to additionally engineer the strain accordingly.

Additional examples on the use of both metabolic engineering and more general engineering strategies for the optimization of secondary metabolites production in actinobacteria have recently been reviewed elsewhere.[56]

4. Conclusions

The expansions and improvements witnessed within the fields of genome and compound mining, omics, and molecular cloning techniques are paving the way for systems metabolic engineering to harness the production potential of streptomycetes. Systematic and global analyses that have been undertaken will continue to improve, and, as a result, will foster faster and better decision making when re-designing a given strain for the optimal production of a target secondary metabolite and discovery of new secondary metabolites. However, it should be noted that, despite all these resources available, a certain amount of iterations of the design-build-test-learn cycle are still necessary as demonstrated for other microbial metabolic engineering cases, especially taking into account the complex regulation between primary and secondary metabolism in streptomycetes. Implementation of molecular cloning methods in an automated, high-throughput setting, such as iBioFab,[80] biosensors for detecting expression of secondary metabolite biosynthesis genes,[81] or biosensors for the production of a desired secondary metabolite[82] might further reduce time and efforts needed to optimize the production of secondary metabolites using streptomycetes. It is expected that the tools and strategies of systems metabolic engineering of streptomycetes will advance rapidly to harness full biotechnological potentials of this important class of bacteria.

Abbreviations

antiSMASH, antibiotics and Secondary Metabolites Analysis Shell; ARC, antibiotic-remodeling compound; ARTS, Antibiotics Resistant Target Seeker; BGC, biosynthetic gene cluster; Cas9, CRISPR-associated protein; CDA, calcium-dependent antibiotic; ChIP, Chromatin ImmunoPrecipitation; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; dRNA-seq, differential RNA sequencing; EFMA, Elementary Flux Mode Analysis; FBA, flux balance analysis; FSEOF, Flux Scanning based on Enforced Objective Flux; GARLIC, Global Alignment for natuRaL-products chemInformatiCs; GC, gas chromatography; GCF, Gene Cluster Family; GNPS, Global Natural Products Social Molecular Networking; GRAPE, Generalized Retrobiosynthetic Assembly Prediction Engine; GSMM, genome-scale metabolic models; HR, homologous recombination; IMG-ABC, Integrated Microbial Genomes Atlas of Biosynthetic gene Clusters; LC-MS/MS, liquid chromatography tandem mass spectrometry; MFA, metabolic flux analysis; MIBiG, Minimum Information on Biosynthetic Gene Clusters; MOMA, Minimization Of Flux Adjustment; NHEJ, non-homologous end joining; NMR, Nuclear magnetic resonance; NRP, non-ribosomal peptide; PAM, protospacer adjacent motif; PI, pristinamycin I; PII, pristinamycin II; PK, polyketide; PRISM, Prediction Informatics for Secondary Metabolomes; Ribo-seq, ribosome profiling; sgRNA, single guide RNA; SMBP, The Secondary Metabolite Bioinformatics Portal; ssRNA-seq, strand-specific RNA sequencing; TFR, TetR family transcriptional regulator; WGCNA, weighted correlation network analysis.

Acknowledgements

This work was funded by a grant from the Novo Nordisk Foundation. H.U. K and S.Y.L. are also supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries (NRF-2012M1A2A2026556 and NRF-2012M1A2A2026557) from the Ministry of Science and ICT through the National Research Foundation (NRF) of Korea.

Conflict of Interest

The authors declare no financial or commercial conflict of interest.

Keywords

biosynthetic gene clusters, genome mining, secondary metabolites, streptomycetes, systems metabolic engineering

Received: September 12, 2017
Revised: October 20, 2017
Published online:

[1] A. L. Demain, S. Sanchez, *J. Antibiot.* **2009**, *62*, 1.
[2] J. Bérdy, *J. Antibiot.* **2005**, *58*, 1.
[3] S. D. Bentley, K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.-A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, D. A. Hopwood, *Nature* **2002**, *417*, 6885.
[4] H. Ikeda, J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, S. Ōmura, *Nat. Biotechnol.* **2003**, *21*, 5.
[5] F. Barona-Gómez, U. Wong, A. E. Giannakopulos, P. J. Derrick, G. L. Challis, *J. Am. Chem. Soc.* **2004**, *126*, 50.
[6] G. L. Challis, J. Ravel, *FEMS Microbiol. Lett.* **2000**, *187*, 2.
[7] S. Lautru, R. J. Deeth, L. M. Bailey, G. L. Challis, *Nat. Chem. Biol.* **2005**, *1*, 5.
[8] B. Zhao, X. Lin, L. Lei, D. C. Lamb, S. L. Kelly, M. R. Waterman, D. E. Cane, *J. Biol. Chem.* **2008**, *283*, 13.
[9] J. P. Gomez-Escribano, L. Song, D. J. Fox, V. Yeo, M. J. Bibb, G. L. Challis, *Chem. Sci.* **2012**, *3*, 9.
[10] M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, R. Breitling, *Nucleic Acids Res.* **2011**, *39*, Web Server issue.
[11] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Müller, W. Wohlleben, R. Breitling, E. Takano, M. H. Medema, *Nucleic Acids Res.* **2015**, *43*, W1.
[12] K. Blin, T. Wolf, M. G. Chevrette, X. Lu, C. J. Schwalen, S. A. Kautsar, H. G. Suarez Duran, E. L. C. de los Santos, H. U. Kim, M. Nave, J. S. Dickschat, D. A. Mitchell, E. Shelest, R. Breitling, E. Takano, S. Y. Lee, T. Weber, M. H. Medema, *Nucleic Acids Res.* **2017**, *45*, W1.
[13] M. A. Skinnider, C. A. Dejong, P. N. Rees, C. W. Johnston, H. Li, A. L. H. Webster, M. A. Wyatt, N. A. Magarvey, *Nucleic Acids Res.* **2015**, *43*, 20.
[14] M. A. Skinnider, N. J. Merwin, C. W. Johnston, N. A. Magarvey, *Nucleic Acids Res.* **2017**, *45*, W1.
[15] K.-S. Hwang, H. U. Kim, P. Charusanti, B. Ø. Palsson, S. Y. Lee, *Biotechnol. Adv.* **2014**, *32*, 2.
[16] G. P. van Wezel, K. J. McDowall, *Nat. Prod. Rep.* **2011**, *28*, 7.
[17] H. U. Kim, P. Charusanti, S. Y. Lee, T. Weber, *Nat. Prod. Rep.* **2016**, *33*, 8.
[18] L. B. Pickens, Y. Tang, Y.-H. Chooi, *Annu. Rev. Chem. Biomol. Eng.* **2011**, 2.
[19] S. Y. Lee, H. U. Kim, *Nat. Biotechnol.* **2015**, *33*, 10.
[20] M. H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J. B. Biggins, K. Blin, I. de Bruijn, Y. H. Chooi, J. Claesen, R. C. Coates, P. Cruz-Morales, S. Duddela, S. Düsterhus, D. J. Edwards, D. P. Fewer, N. Garg, C. Geiger, J. P. Gomez-Escribano, A. Greule, M. Hadjithomas, A. S. Haines, E. J. N. Helfrich, M. L. Hillwig, K. Ishida, A. C. Jones, C. S. Jones, K. Jungmann, C. Kegler, H. U. Kim, P. Kötter, D. Krug, J. Masschelein, A. V. Melnik, S. M. Mantovani, E. A. Monroe, M. Moore, N. Moss, H.-W. Nützmann, G. Pan, A. Pati, D. Petras, F. J. Reen, F. Rosconi, Z. Rui, Z. Tian, N. J. Tobias, Y. Tsunematsu, P. Wiemann, E. Wyckoff, X. Yan, G. Yim, F. Yu, Y. Xie, B. Aigle, A. K. Apel, C. J. Balibar, E. P. Balskus, F. Barona-Gómez, A. Bechthold, H. B. Bode, R. Borriss, S. F. Brady, A. A. Brakhage, P. Caffrey, Y.-Q. Cheng, J. Clardy, R. J. Cox, R. De Mot, S. Donadio, M. S. Donia, W. A. van der Donk, P. C. Dorrestein, S. Doyle, A. J. M. Driessen, M. Ehling-Schulz, K.-D. Entian, M. A. Fischbach, L. Gerwick, W. H. Gerwick, H. Gross, B. Gust, C. Hertweck, M. Höfte, S. E. Jensen, J. Ju, L. Katz, L. Kaysser, J. L. Klassen, N. P. Keller, J. Kormanec, O. P. Kuipers, T. Kuzuyama, N. C. Kyrpides, H.-J. Kwon, S. Lautru, R. Lavigne, C. Y. Lee, B. Linquan, X. Liu, W. Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Méndez, M. Metsä-Ketelä, J. Micklefield, D. A. Mitchell, B. S. Moore, L. M. Moreira, R. Müller, B. A. Neilan, M. Nett, J. Nielsen, F. O'Gara, H. Oikawa, A. Osbourn, M. S. Osburne, B. Ostash, S. M. Payne, J.-L. Pernodet, M. Petricek, J. Piel, O. Ploux, J. M. Raaijmakers, J. A. Salas, E. K. Schmitt, B. Scott, R. F. Seipke, B. Shen, D. H. Sherman, K. Sivonen, M. J. Smanski, M. Sosio, E. Stegmann, R. D. Süssmuth, K. Tahlan, C. M. Thomas, Y. Tang, A. W. Truman, M. Viaud, J. D. Walton, C. T. Walsh, T. Weber, G. P. van Wezel, B. Wilkinson, J. M. Willey, W. Wohlleben, G. D. Wright, N. Ziemert, C. Zhang, S. B. Zotchev, R. Breitling, E. Takano, *Nat. Chem. Biol.* **2015**, *11*, 9.
[21] M. Hadjithomas, I.-M. A. Chen, K. Chu, A. Ratner, K. Palaniappan, E. Szeto, J. Huang, T. B. K. Reddy, P. Cimermancic, M. A. Fischbach, N. N. Ivanova, V. M. Markowitz, N. C. Kyrpides, A. Pati, *MBio* **2015**, *6*, 4.
[22] K. Blin, M. H. Medema, R. Kottmann, S. Y. Lee, T. Weber, *Nucleic Acids Res.* **2017**, *45*, D1.
[23] A. Flissi, Y. Dufresne, J. Michalik, L. Tonon, S. Janot, L. Noé, P. Jacques, V. Leclère, M. Pupin, *Nucleic Acids Res.* **2016**, *44*, D1.
[24] C. W. Johnston, N. A. Magarvey, *Nat. Chem. Biol.* **2015**, *11*, 3.
[25] D. Klementz, K. Döring, X. Lucas, K. K. Telukunta, A. Erxleben, D. Deubel, A. Erber, I. Santillana, O. S. Thomas, A. Bechthold, S. Günther, *Nucleic Acids Res.* **2015**, *44*, D1.
[26] P. Cruz-Morales, J. F. Kopp, C. Martínez-Guerrero, L. A. Yáñez-Guerra, N. Selem-Mojica, H. Ramos-Aboites, J. Feldmann, F. Barona-Gómez, *Genome Biol. Evol.* **2016**, *8*, 6.
[27] M. Alanjary, B. Kronmiller, M. Adamek, K. Blin, T. Weber, D. Huson, B. Philmus, N. Ziemert, *Nucleic Acids Res.* **2017**, *45*, W1.
[28] J. R. Doroghazi, J. C. Albright, A. W. Goering, K.-S. Ju, R. R. Haines, K. A. Tchalukov, D. P. Labeda, N. L. Kelleher, W. W. Metcalf, *Nat. Chem. Biol.* **2014**, *10*, 11.
[29] C. Wu, H. K. Kim, G. P. van Wezel, Y. H. Choi, *Drug Discov. Today Technol.* **2015**, *13*, 11.

[30] C. A. Dejong, G. M. Chen, H. Li, C. W. Johnston, M. R. Edwards, P. N. Rees, M. A. Skinnider, A. L. H. Webster, N. A. Magarvey, *Nat. Chem. Biol*. **2016**, *12*, 12.
[31] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. P Boya, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, N. Bandeira, *Nat. Biotechnol*. **2016**, *34*, 8.
[32] T. Weber, *Int. J. Med. Microbiol*. **2014**, *304*, 3.
[33] M. H. Medema, M. A. Fischbach, *Nat. Chem. Biol*. **2015**, *11*, 9.
[34] N. Ziemert, M. Alanjary, T. Weber, *Nat. Prod. Rep*. **2016**, *33*, 8.
[35] T. Weber, K. Blin, H. U. Kim, *Synth. Syst Biotechnol*. **2016**, *1*, 2.
[36] A. K. Chaudhary, D. Dhakal, J. K. Sohng, *Biomed Res. Int*. **2013**, *3*, 8.
[37] K. Nieselt, F. Battke, A. Herbig, P. Bruheim, A. Wentzel, M. Ø. Jakobsen, H. Sletta, M. T. Alam, M. E. Merlo, J. Moore, W. A. M. Omara, E. R. Morrissey, M. A. Juarez-Hermosillo, A. Rodríguez-García, M. Nentwich, L. Thomas, M. Iqbal, R. Legaie, W. H. Gaze, G. L. Challis, R. C. Jansen, L. Dijkhuizen, D. A. Rand, D. L. Wild, M. Bonin, J. Reuther, W. Wohlleben, M. C. M. Smith, N. J. Burroughs, J. F. Martín, D. A. Hodgson, E. Takano, R. Breitling, T. E. Ellingsen, E. M. H. Wellington, *BMC Genomics* **2010**, *11*, 10.
[38] Y. Jeong, J.-N. Kim, M. W. Kim, G. Bucca, S. Cho, Y. J. Yoon, B.-G. Kim, J.-H. Roe, S. C. Kim, C. P. Smith, B.-K. Cho, *Nat. Commun*. **2016**, *7*, 11605.
[39] A. Manteca, J. Sanchez, H. R. Jung, V. Schwämmle, O. N. Jensen, *Mol. Cell. Proteomics* **2010**, *9*, 7.
[40] M. M. Zhang, Y. Wang, E. L. Ang, H. Zhao, *Nat. Prod. Rep*. **2016**, *33*, 8.
[41] B. Gust, G. Chandra, D. Jakimowicz, T. Yuqing, C. J. Bruton, K. F. Chater, *Adv. Appl. Microbiol*. **2004**, *54*, 107.
[42] T. Siegl, L. Petzke, E. Welle, A. Luzhetskyy, *Appl. Microbiol. Biotechnol*. **2010**, *87*, 4.
[43] L. T. Fernández-Martínez, M. J. Bibb, *Sci Rep*. **2014**, *4*, 7100.
[44] G. Muth, B. Nußbaumer, W. Wohlleben, A. Pühler, *MGG Mol. Gen. Genet*. **1989**, *219*, 3.
[45] R. E. Cobb, Y. Wang, H. Zhao, *ACS Synth. Biol*. **2015**, *4*, 6.
[46] H. Huang, G. Zheng, W. Jiang, H. Hu, Y. Lu, *Acta Biochim. Biophys. Sin. (Shanghai)* **2015**, *47*, 4.
[47] Y. Tong, P. Charusanti, L. Zhang, T. Weber, S. Y. Lee, *ACS Synth. Biol*. **2015**, *4*, 9.
[48] M. M. Zhang, F. T. Wong, Y. Wang, S. Luo, Y. H. Lim, E. Heng, W. L. Yeo, R. E. Cobb, B. Enghiad, E. L. Ang, H. Zhao, *Nat. Chem. Biol*. **2017**, *13*, 6.
[49] Z. Qin, J. T. Munnoch, R. Devine, N. A. Holmes, R. F. Seipke, K. A. Wilkinson, B. Wilkinson, M. I. Hutchings, *Chem. Sci*. **2017**, *8*, 4.
[50] T. C. McLean, P. A. Hoskisson, R. F. Seipke, *mSphere* **2016**, *1*, 6.
[51] T. Wolf, T. Gren, E. Thieme, D. Wibberg, T. Zemke, A. Pühler, J. Kalinowski, *J. Biotechnol*. **2016**, *231*, 122.
[52] C. Olano, I. García, A. González, M. Rodriguez, D. Rozas, J. Rubio, M. Sánchez-Hidalgo, A. F. Braña, C. Méndez, J. A. Salas, *Microb Biotechnol*. **2014**, *7*, 3.
[53] Y. Luo, H. Huang, J. Liang, M. Wang, L. Lu, Z. Shao, R. E. Cobb, H. Zhao, *Nat. Commun*. **2013**, *4*, 2894.
[54] T. Siegl, B. Tokovenko, M. Myronovskyi, A. Luzhetskyy, *Metab. Eng*. **2013**, *19*, 98.
[55] Y. Luo, L. Zhang, K. W. Barton, H. Zhao, *ACS Synth. Biol*. **2015**, *4*, 9.
[56] O. Bilyk, A. Luzhetskyy, *Curr. Opin. Biotechnol*. **2016**, *42*, 98.
[57] Y. Luo, B. Enghiad, H. Zhao, *Nat. Prod. Rep*. **2016**, *33*, 2.
[58] R. H. Baltz, *J. Ind. Microbiol. Biotechnol*. **2016**, *43*, 2.
[59] J. P. Gomez-Escribano, M. J. Bibb, *Microb Biotechnol*. **2011**, *4*, 2.
[60] M. Zhou, X. Jing, P. Xie, W. Chen, T. Wang, H. Xia, Z. Qin, *FEMS Microbiol. Lett*. **2012**, *333*, 2.
[61] M. R. Seyedsayamdost, *Proc. Natl. Acad. Sci*. **2014**, *111*, 20.
[62] U. R. Abdelmohsen, T. Grkovic, S. Balasubramanian, M. S. Kamel, R. J. Quinn, U. Hentschel, *Biotechnol. Adv*. **2015**, *33*, 6.
[63] Y. Tanaka, T. Hosaka, K. Ochi, *J. Antibiot*. **2010**, *63*, 8.
[64] A. Craney, C. Ozimok, S. M. Pimentel-Elardo, A. Capretta, J. R. Nodwell, *Chem. Biol*. **2012**, *19*, 8.
[65] M. N. Thaker, W. Wang, P. Spanogiannopoulos, N. Waglechner, A. M. King, R. Medina, G. D. Wright, *Nat. Biotechnol*. **2013**, *31*, 10.
[66] S. M. Pimentel-Elardo, D. Sørensen, L. Ho, M. Ziko, S. A. Bueler, S. Lu, J. Tao, A. Moser, R. Lee, D. Agard, G. Fairn, J. L. Rubinstein, B. K. Shoichet, J. R. Nodwell, *ACS Chem. Biol*. **2015**, *10*, 11.
[67] C. Wang, J. Liu, H. Liu, S. Liang, J. Wen, *J. Ind. Microbiol. Biotechnol*. **2017**, *44*, 11.
[68] C. Wang, J. Liu, H. Liu, J. Wang, J. Wen, *Biochem. Eng. J*. **2017**, *123*, 45.
[69] J. Wang, H. Liu, D. Huang, L. Jin, C. Wang, J. Wen, *Appl. Microbiol. Biotechnol*. **2017**, *101*, 6.
[70] H. Qi, X. Xin, S. Li, J. Wen, Y. Chen, X. Jia, *Biotechnol Bioprocess Eng*. **2012**, *17*, 4.
[71] H. Qi, S. Zhao, J. Wen, Y. Chen, X. Jia, *Biochem. Eng. J*. **2014**, *82*, 124.
[72] H. Qi, M. Lv, K. Song, J. Wen, *Biotechnol. Bioeng*. **2017**, *114*, 5.
[73] J. Chen, M. Liu, X. Liu, J. Miao, C. Fu, H. Gao, R. Müller, Q. Zhang, L. Zhang, *Synth. Syst Biotechnol*. **2016**, *1*, 1.
[74] L. Li, Y. Zhao, L. Ruan, S. Yang, M. Ge, W. Jiang, Y. Lu, *Metab. Eng*. **2015**, *29*, 12.
[75] J. Meng, R. Feng, G. Zheng, M. Ge, Y. Mast, W. Wohlleben, J. Gao, W. Jiang, Y. Lu, *Synth. Syst Biotechnol*. **2017**, *2*, 2.
[76] J. Guo, X. Zhang, S. Luo, F. He, Z. Chen, Y. Wen, J. Li, *PLoS One* **2013**, *8*, 8.
[77] J. Guo, X. Zhang, Z. Chen, Y. Wen, J. Li, *PLoS One* **2014**, *9*, 6.
[78] Y. Mast, T. Weber, M. Gölz, R. Ort-Winklbauer, A. Gondran, W. Wohlleben, E. Schinko, *Microb Biotechnol*. **2010**, *4*, 2.
[79] Y. Mast, J. Guezguez, F. Handel, E. Schinko, *Appl. Environ. Microbiol*. **2015**, *81*, 19.
[80] R. Chao, J. Liang, I. Tasan, T. Si, L. Ju, H. Zhao, *ACS Synth. Biol*. **2017**, *6*, 4.
[81] F. Guo, S. Xiang, L. Li, B. Wang, J. Rajasärkkä, K. Gröndahl-Yli-Hannuksela, G. Ai, M. Metsä-Ketelä, K. Yang, *Metab. Eng*. **2015**, *28*, 134.
[82] Y.-Q. Sun, T. Busche, C. Rückert, C. Paulus, Y. Rebets, R. Novakova, J. Kalinowski, A. Luzhetskyy, J. Kormanec, O. N. Sekurova, S. B. Zotchev, *ACS Synth. Biol*. **2017**, *6*, 6.
[83] I. Borodina, P. Krabben, J. Nielsen, *Genome Res*. **2005**, *15*, 6.

[84] I. Borodina, J. Siebring, J. Zhang, C. P. Smith, G. van Keulen, L. Dijkhuizen, J. Nielsen, *J. Biol. Chem.* **2008**, *283*, 37.
[85] M. Kim, J. Sang Yi, J. Kim, J.-N. Kim, M. W. Kim, B.-G. Kim, *Biotechnol. J.* **2014**, *9*, 9.
[86] D. Huang, X. Jia, J. Wen, G. Wang, G. Yu, Q. Caiyin, Y. Chen, *Appl. Biochem. Biotechnol.* **2011**, *165*, 7.
[87] D. Huang, J. Wen, G. Wang, G. Yu, X. Jia, Y. Chen, *Appl. Microbiol. Biotechnol.* **2012**, *94*, 3.
[88] L. Dang, J. Liu, C. Wang, H. Liu, J. Wen, *J. Ind. Microbiol. Biotechnol.* **2017**, *44*, 2.
[89] G.-Y. Tan, K. Deng, X. Liu, H. Tao, Y. Chang, J. Chen, K. Chen, Z. Sheng, Z. Deng, T. Liu, *ACS Synth. Biol.* **2017**, *6*, 6.

Manuscripts II and III

The CRISPR-Cas9 system

The clustered regularly interspaced short palindromic repeats (CRISPR) – CRISPR-associated nuclease (Cas9) system was first recognised as a part of the adaptive immune response in bacteria and archeae[65–67]. Upon encountering foreign DNA or RNA, the host organism stores short fragments of genetic information, termed protospacers, between identical repeats in so-called CRISPR arrays. Following this adaption phase, the host enters the expression phase in which precursor CRISPR RNA (pre-crRNA) is transcribed and further processed to yield short crRNAs. During the final phase of interference, the crRNAs can bind to the complementary protospacer sequences present on the invading viral or plasmid target at which point the Cas proteins are recruited[66,68] (**Fig. 10**). To date, three different CRISPR-Cas systems (I, II, III) have been characterised[69]. Of these, the CRISPR type II has received special attention, in that it relies only on three distinct components; the crRNA, *trans*-activating crRNA (tracrRNA), and a single Cas9 protein[68,70]. The role of the tracrRNA is dual in that it both catalyses the processing of pre-crRNA into crRNA and is involved in formation of the complex of Cas9 and crRNA:tracrRNA. Upon binding of this complex to the complementary sequence in the target DNA, Cas9 undergoes conformation changes[71], leading to activation of the HNH and RuvC endonuclease domains and with that the introduction of double stranded breaks (DSBs) in the target DNA.

The versatility of the CRISPR-Cas9 system potentiates its use in numerous applications, including strain genotyping, epidemiological studies, strain improvements, and transcriptional modulation[66,72]. Strain improvement in particular has been subject to much progress. The development of a single guide RNA (sgRNA), composed of the crRNA:tracrRNA complex, has enabled genetic modification of a wide range of hosts, including bacteria, yeast, human cell lines, mouse, zebrafish, and others[73]. In addition, the construction of a catalytically “dead” Cas9 (dCas9), mutated in the HNH (D10A) and RuvC (H840A) domains, has facilitated the use of the protein as a transcriptional modulator, in that the dCas9 retains its ability to bind to DNA, however, lacks the endonuclease activity[68].

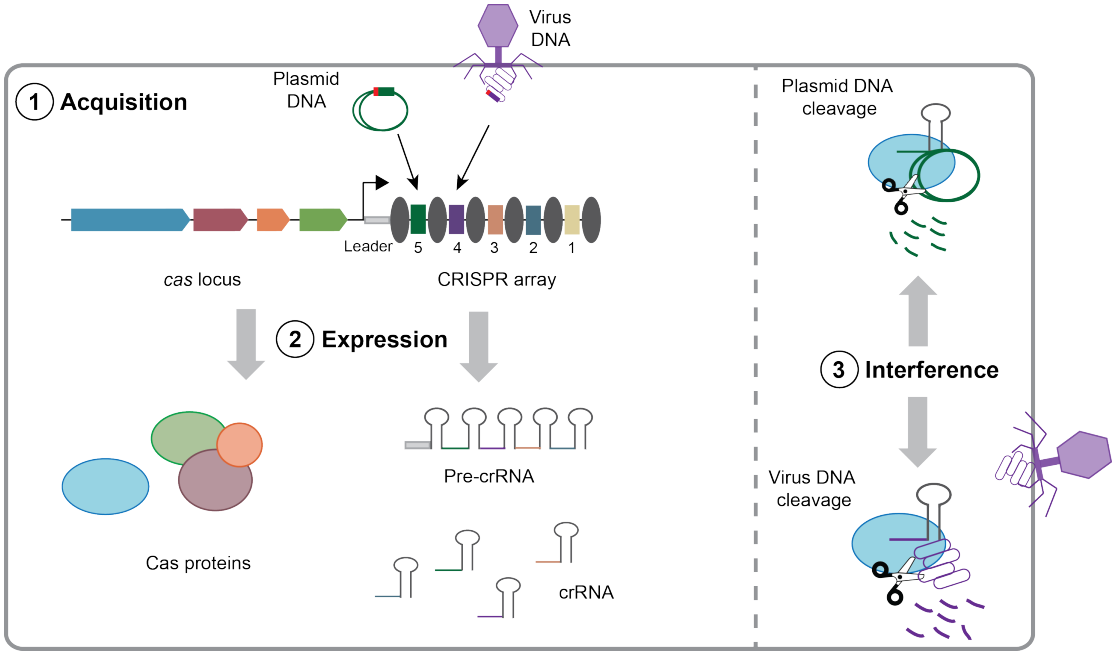


Figure 10 The three stages of CRISPR-Cas adaptive immunity in prokaryotes including acquisition (1), expression (2), and interference (3). In (1), the bacteria is challenged with viral or plasmid DNA resulting in acquisition of unique protospacers (coloured squared with PAM (red) upstream), which are stored between the repetitive sequences (grey ellipses) in the CRISPR array. During (2), the leader region (light grey box) of the CRISPR array is transcribed into pre-crRNA, which is further processed by Cas proteins into crRNAs, containing single spacers and partial repeats. When challenged with viral or plasmid DNA (3) with strong similarity to the acquired protospacers, the incoming DNA is cleaved by the Cas proteins. This general scheme is based on the well-studied and relatively simple CRISPR-Cas system of *Streptococcus thermophiles*[65,74]. Cas, CRISPR-associated; CRISPR, clustered regularly interspaced short palindromic repeats; crRNA, CRISPR RNA; Pre-crRNA, precursor CRISPR RNA; PAM, protospacer-adjacent motif.

The development of a CRISPR-Cas9 technology suitable for genome editing in actinomycetes[75–77] has opened up new ways of engineering the strains for production of secondary metabolites as evident from recent publications[78–81].

To promote the use of the CRISPR-Cas9 system for scientist working in the field of streptomycetes, our group has published a "CRISPR-Cas9 Toolkit for Actinomycete Genome Editing". The method, which is a chapter in the Springer Protocol series "Synthetic Metabolic Pathways", is presented as **manuscript II** in this dissertation.

To further expand the applicability of the USER-CRISPR-Cas9 cloning platform, an optimised USER cloning protocol, compatible with a high-throughput setting, for vector construction was developed. The development and proof-of-concept (PoC) of the optimised workflow are presented in **manuscript III** entitled "An improved USER-CRISPR-Cas9 platform for high-throughput genome editing in streptomycetes".

Manuscript II

CRISPR-Cas9 Toolkit for Actinomycete Genome Editing

Yaojun Tong[1], Helene Lunde Robertsen[1], Kai Blin[1], Tilmann Weber[1], and Sang Yup Lee[1,2]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark.

[2]Department of Chemical and Biomolecular Engineering and BioInformatics Research Center, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea.

(Accepted chapter included in "Synthetic Metabolic Pathways: Methods and Protocols", part of the Springer Book series "Methods in Molecular Biology")
(Published online November 24th 2017)

Chapter 11

CRISPR-Cas9 Toolkit for Actinomycete Genome Editing

Yaojun Tong, Helene Lunde Robertsen, Kai Blin, Tilmann Weber, and Sang Yup Lee

Abstract

Bacteria of the order Actinomycetales are one of the most important sources of bioactive natural products, which are the source of many drugs. However, many of them still lack efficient genome editing methods, some strains even cannot be manipulated at all. This restricts systematic metabolic engineering approaches for boosting known and discovering novel natural products. In order to facilitate the genome editing for actinomycetes, we developed a CRISPR-Cas9 toolkit with high efficiency for actinomyces genome editing. This basic toolkit includes a software for spacer (sgRNA) identification, a system for in-frame gene/gene cluster knockout, a system for gene loss-of-function study, a system for generating a random size deletion library, and a system for gene knockdown. For the latter, a uracil-specific excision reagent (USER) cloning technology was adapted to simplify the CRISPR vector construction process. The application of this toolkit was successfully demonstrated by perturbation of genomes of *Streptomyces coelicolor* A3(2) and *Streptomyces collinus* Tü 365. The CRISPR-Cas9 toolkit and related protocol described here can be widely used for metabolic engineering of actinomycetes.

Key words CRISPR-Cas9, CRISPRi, Uracil-specific excision reagent (USER) cloning, Synthetic biology, Actinomycetes, Genome editing, Double-strand break (DSB), Homology directed repair (HDR), Non-homologous end joining (NHEJ)

1 Introduction

1.1 Actinomycetes

Actinomycetes are Gram-positive bacteria with high GC content genomes, belonging to the order of Actinomycetales. They are well known for their ability to produce medically and industrially relevant secondary metabolites (natural products) [1–3], including, but not limited to antibiotics, herbicides, chemotherapeutics, and immunosuppressants, such as vancomycin, bialaphos, doxorubicin, and rapamycin, respectively. However, after being studied over half-century, it becomes more and more challenging to find novel secondary metabolites with meaningful properties by traditional

The authors have filed a patent (EP15160126.7) on the actinomycete CRISPR toolkit.

Michael Krogh Jensen and Jay D. Keasling (eds.), *Synthetic Metabolic Pathways: Methods and Protocols*, Methods in Molecular Biology, vol. 1671, https://doi.org/10.1007/978-1-4939-7295-1_11,

methods. However, modern genome mining techniques [4–7] have revealed that those bacteria still possess a huge unexploited potential to produce secondary metabolites with novel structures [8]. Unfortunately, in comparison with model organisms like *E. coli* and *S. cerevisiae*, there are only few genetic manipulation tools available for actinomycetes. In addition, the high GC content (sometimes >72%) impedes genetic manipulation even if actinomycete DNA is manipulated in other hosts like *E. coli*. With the help of the recently developed CRISPR-Cas9 technology, we now have more tools to address and overcome these challenges for efficient genetic manipulation of actinomycetes.

1.2 CRISPR-Cas9

The modules of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated (Cas) proteins are present in most archaea and many bacteria as adaptive immune systems for defense against foreign DNA [9–11] or RNA [12]. Based on the number of Cas proteins involved, CRISPR-Cas systems can be divided into "class 1" and "class 2." Class 1 systems have multiple Cas9 proteins, while class 2 systems only need one single Cas protein, for instance, type II CRISPR system, also known as CRISPR-Cas9 system [13]. The currently well-studied and widely used CRISPR-Cas9 system is originally from *Streptococcus pyogenes*. The target sequences of the Cas9 endonuclease are defined in the CRISPR loci containing short repeats separated by "spacer" sequences that exactly match the sequences of the targeted foreign genetic element. Introducing double-strand breaks (DSBs) into these DNAs offers adaptive immunity against foreign genetic elements [9, 14–17]. In the native CRISPR-Cas9 system, the spacer sequence of the CRISPR array transcribes to a CRISPR RNA (crRNA). Subsequently, an associated trans-activating CRISPR RNA (tracrRNA) hybridizes with the crRNA, forming an RNA duplex, which is cleaved and further processed by endogenous RNase III and possibly other, yet unknown nucleases [18]. The crRNA-tracrRNA duplex, which was later artificially designed as a chimera named "single guide RNA" (sgRNA) [17], interacts with Cas9 to form a complex, then scans the foreign genetic elements for the presence of trinucleotide protospacer adjacent motifs (PAMs). When this complex finds a PAM that has a 5′ sequence (normally around 20 nt) complementary to the spacer sequence in the crRNA-tracrRNA duplex, it binds to this position and then triggers the conformational change of Cas9 to activate the HNH and RuvC endonuclease domains [19, 20], which causes DNA double-strand break (DSB). The DNA DSB of a chromosome is lethal, and cells can only survive if the lesion is repaired. The two major routes for DNA repair are (1) non-homologous end joining (NHEJ), in which no editing template is needed, and (2) homology-directed repair (HDR), in which an editing template for homologous recombination is needed [21, 22] (Fig. 1).

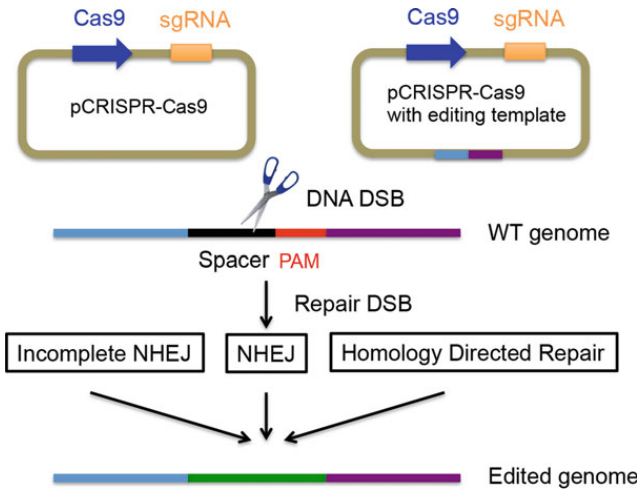


Fig. 1 CRISPR-Cas9 schematics. CRISPR-Cas9 genome editing working model

Using the feature of introducing DNA DSBs, which then get repaired by cellular mechanisms such as NHEJ or HDR, allows the use of CRISPR-Cas9 for genome editing in many different organisms, from *E. coli* to human cells [23–31]. NHEJ is a complicated process that involves several proteins. It has been best described in eukaryotic cells to repair the DNA DSB in an error-prone manner [32, 33]; however, it is also found in prokaryotes [34, 35]. Interestingly, in *S. coelicolor* we found that the NHEJ in contrast to other bacteria, e.g., *Mycobacterium tuberculosis* [34], is missing a DNA ligase function. In this strain, the activity of this enzyme can be partially restored by other yet unknown enzymes, albeit with a lower efficiency. In *S. coelicolor* and other streptomycetes lacking the ligase gene, the native "incomplete" NHEJ repair pathway usually leads to larger deletions around the site of the DSB. This phenomenon can be exploited to trigger deletions between the nearest two essential genes, thus leading to "random size deletion libraries" [36, 37].

A study of *S. pyogenes* Cas9 nuclease domain revealed that mutating the HNH and RuvC domains (D10A and H840A) resulted in a catalytically dead Cas9 (dCas9) variant that does not have endonuclease activity, but could still form a complex with sgRNA and efficiently bind to the target DNA [17]. This effect can be used to sequence-specifically interfere with transcription and thus control gene expression. In analogy to eukaryotic RNA interference (RNAi), this system was named as CRISPRi [38].

1.3 USER Cloning

Construction of CRISPR-Cas9 vectors using ligation-based approach is still relatively time consuming, and is difficult to be implemented in high-throughput and automation settings.

Nowadays, PCR-based cloning is a commonly used method for de novo gene assembly in metabolic engineering [39]. USER friendly cloning is one of those modern cloning methods, and it represents an alternative to conventional ligation-based cloning in that it allows for simultaneous scarless assembly of multiple PCR products into USER-compatible vectors (Fig. 2). This allows for easy and versatile vector construction [40]. We introduced USER friendly cloning to facilitate CRISPR-Cas9 vectors construction, as well as to meet the demands for further high-throughput and automated genome editing purposes.

USER assembly relies on the generation of complementary overhangs in the PCR products and destination vector and can be divided into three distinct steps. First, genes of interest (GOIs) are PCR amplified with primers containing between 7 and 12 nucleotides overhangs flanked by uracil bases (dU) [41]. The directional assembly of the PCR fragments is facilitated through the overhangs, which are designed either manually or using an online tool such as AMUSER 1.0 (at http://www.cbs.dtu.dk/services/AMUSER/) [42]. In addition, a proofreading DNA polymerase such as PfuX7 (Norholm, *see* [43]) or the commercially available Phusion U Hot Start DNA Polymerase (Thermo Fisher Scientific, Waltham, US) is required for the recognition of the uracil bases and incorporation of adenosine residues on the complementary strand; second, the destination vector is linearized and with that complementary overhangs generated using a combination of a restriction and a nicking enzyme. The enzymes required for linearization and generation of single-stranded overhangs depend on the USER cassette in the destination vector. Examples of USER cassettes include the PacI/Nt.BbvCI, AsiSI/Nb.BsmI, and AsiSI/Nb.BtsI cassettes [44]; in the third step, PCR fragments are assembled in the linearized vector by means of the USER™ kit (New England Biolabs) that contains a mixture of the *E. coli* uracil DNA glycosylase and DNA glycosylase-lyase endonuclease VIII, both of which recognize and remove uracil bases. Following uracil excision, the reaction is kept at the melting temperature of the single-stranded overhangs for several minutes to facilitate the assembly of the PCR fragments in the destination vector [45].

During the past 2 years, independent laboratories have established modular and efficient genetic manipulation tools for streptomycetes based on CRISPR-Cas9. These tools significantly facilitated the processes of gene/gene cluster deletion, point mutagenesis, gene replacement, as well as repression of gene transcription in *Streptomyces* [36, 46–48]. In this chapter, we describe protocols using the toolkit developed in our lab [36, 37] and a workflow combining of in silico primer design for sgRNA construction, USER-based cloning, and CRISPRi.

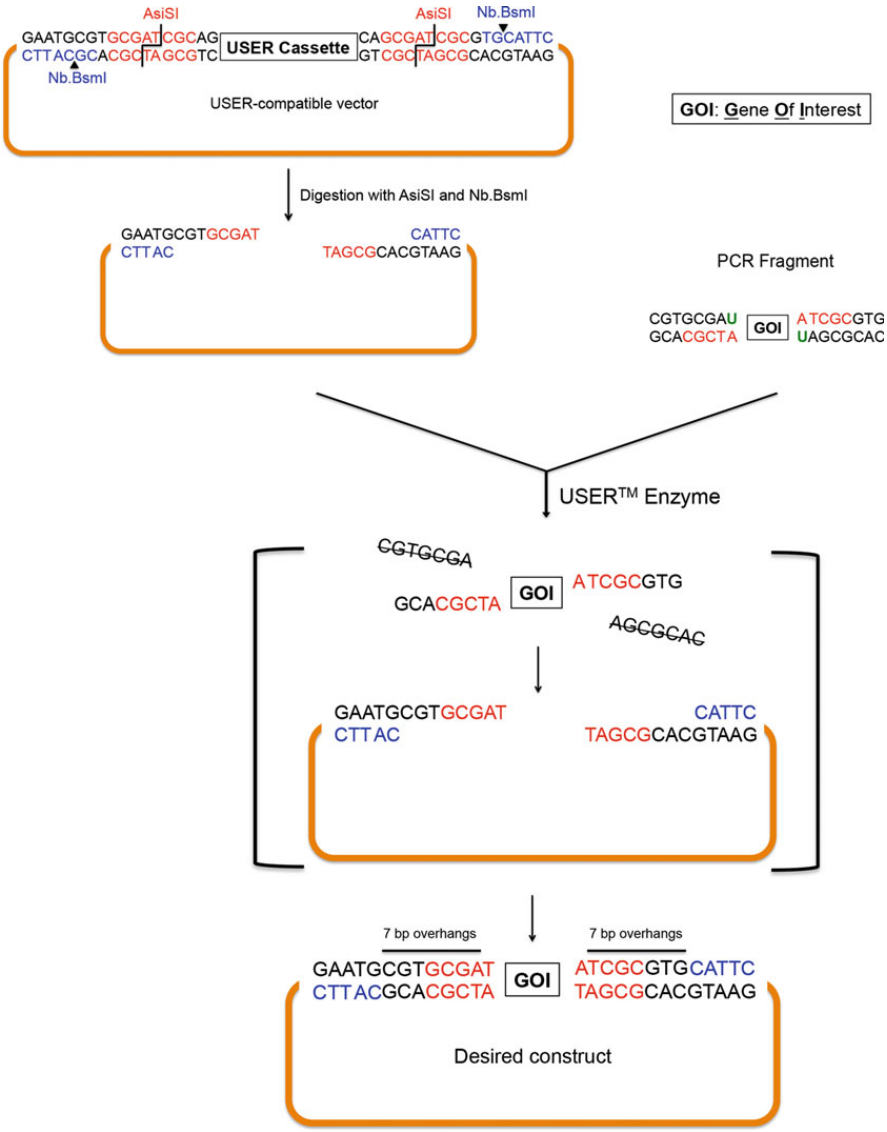


Fig. 2 USER cloning schematics. Schematic overview of the steps involved in USER assembly of PCR-generated fragments in a linearized USER-compatible destination vector

2 Materials

Milli-Q water (18.2 MΩ cm at 25 °C) is used for preparation of all media and solutions. All kits and reagents are used according to the manufacturer's instructions, unless the modifications are indicated. Diligently follow all waste disposal regulations when disposing waste materials.

2.1 Strains

One Shot® ccdB Survival™ 2 T1^{R} chemically competent cells (Thermo Fisher Scientific) are used for the construction of the USER-compatible vectors. Chemically competent *E. coli* cells, e.g., NEB5-α (New England Biolabs), and One Shot® Mach1™ T1 Phage-Resistant (Thermo Fisher Scientific) are used for routine cloning. Non-methylating *E. coli* ET12567/pUZ8002 [49] is used for conjugation. *Streptomyces coelicolor* A3(2), and *Streptomyces collinus* Tü365 are used as example strains in these protocols.

2.2 Plasmids

Plasmids pGM1190 [50], pGM1190-Cas9, pCRISPR-Cas9, pCRISPR-dCas9, pCRISPR-Cas9-ScaligD [36], and pCRISPR-USER-(d)Cas9 are used in the following protocols. All oligonucleotides and gBlocks are purchased from Integrated DNA Technologies (IDT).

2.3 Media

All components for media preparation are purchased from Sigma-Aldrich, unless indicated otherwise.

SOC medium (20 g/L Tryptone, 5 g/L Yeast extract, 4.8 g/L $MgSO_4$, 3.603 g/L Dextrose, 0.5 g/L NaCl, 0.186 g/L KCl), LB medium (10 g/L Tryptone, 5 g/L Yeast extract, 5 g/L NaCl, 20 g/L Agar is added for solidification), ISP2 medium (Yeast extract 4 g/L, Malt extract 10 g/L, Dextrose 4 g/L, 20 g/L Agar is added for solidification), and Soya Flour Mannitol agar (MS, or SFM, or Cullum agar) (20 g/L Mannitol, 20 g/L Soya flour with low fat (W. Schoenenberger GmbH & Co.), 20 g/L Agar) supplemented with 10 mM $MgCl_2$. Appropriate antibiotics are added to the media when needed. Their working concentrations are: apramycin, 50 μg/mL; nalidixic acid, 50 μg/mL; thiostrepton, 1 μg/mL; kanamycin, 25 μg/mL; and chloramphenicol, 25 μg/mL.

2.4 Reagents and Kits

Phusion Hot Start II DNA Polymerase (2 U/μL), DreamTaq Green PCR Master Mix (2×), PCR Master Mix (2×), PfuX7 DNA polymerase [43], Phusion U Hot Start DNA Polymerase (2 U/μL), T4 DNA Ligase (1 U/μL), GeneJET Plasmid Miniprep Kit, CloneJET PCR Cloning Kit, GeneJET PCR Purification Kit, and all restriction enzymes are purchased from Thermo Fisher Scientific; USER™ Enzyme, and Gibson Assembly® Cloning Kit are purchased from New England Biolabs. Blood & Cell Culture

DNA Kit is from Qiagen. NucleoSpin® Gel Clean-up kit is from Macherey-Nagel.

2.5 Equipment

NanoDrop 2000 (Thermo Scientific) is used to measure DNA concentrations, and Concentrator plus (Eppendorf) is used for concentrating DNA solutions.

3 Methods

All the procedures are carried out at room temperature unless otherwise specified, all DNAs are eluted by nuclease-free water (pH 8). The annealing temperature (T_a) is calculated with the Thermo Fisher Scientific T_m calculator: http://www.thermofisher.com/dk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html.

The plasmids within our CRISPR-Cas9 toolkit are based on one single temperature sensitive vector pGM1190, which is a derivate of the replicon pSG5 [50]. All necessary elements are integrated into one single construct, where Cas9/dCas9 is under control of the thiostrepton inducible *tipA* promoter, while the sgRNA is driven by a constitutive *ermE** promoter. The editing template, NHEJ missing component(s), and other element(s) can be inserted via the singular StuI site of the vector [36] (*see* **Note 1**).

3.1 Identification of Suitable 20 nt Spacers with CRISPy-Web

For successful CRISPR/Cas9 experiments it is essential to define good 20 nt spacer sequences within the desired target region of the genome. One prerequisite for these 20 nt spacers is that they match the 20 nt upstream of a PAM close to the desired target. In addition, it is important to avoid off-target effects: If the same or a very similar 20 nt spacer sequence is found close to a PAM elsewhere on the genome, Cas9 will introduce a DSB there as well. To reduce the probability of unwanted side effects even further, spacers that match many other spacers with a mismatch of one or two bases should be avoided.

For this reason, it is recommended to use computational tools to design the guide RNAs. While many programs exist for designing sgRNAs for model organisms, only few tools can be used with user-supplied genomes [51]. The CRISPy-web tool [52] assists researchers in this task by identifying appropriate 20 nt spacer regions for sgRNAs in any user-supplied microbial genome sequence. CRISPy-web is available at http://crispy.secondarymetabolites.org/.

1. To run CRISPy-web for a genome region of interest, a GenBank-formatted file can be uploaded by clicking the "Browse" button and selecting the appropriate file on the

start page of CRISPy-web. Alternatively, CRISPy-web supports directly using the results from the antiSMASH secondary metabolite biosynthetic gene cluster mining platform [4, 5], by simply selecting "Get sequences from antiSMASH" and providing the antiSMASH job id that is included in the antiSMASH result email. The search is started by clicking the "Start" button.

2. Once the sequence has been uploaded, the user has to select a target region to scan for suitable spacers. The region selection page gives a short summary of the uploaded genome and a search field below the summary can be used to specify the target region.

   Regions can be selected by entering a range of nucleotide coordinates (like 12,345–67,890), a gene name, or a locus tag from the annotated genome.

   The search field will autocomplete for gene and cluster name as well as locus tags. If the data were directly transferred from antiSMASH and secondary metabolite biosynthetic gene cluster has been identified, these are also shown in a summary table below the search field.

   A help screen for the syntax is displayed by clicking on "Usage hints."
3. Once a target region is selected, a click on the "Find targets" button starts the actual scan for spacers. Depending on the size of the selected region and the overall genome size, this process usually takes between few—for small—to around 15 min for large genomes.
4. Once the scan has completed, the user will be presented with a page showing an overview of the scanned region. In this overview, genes are displayed as gray arrows and potential 20 nt spacers are indicated as red boxes. It is possible to zoom in to a specific gene by clicking the gene arrow and selecting "Show results for this gene only" from the pop-up.
5. Potential spacers are displayed sorted by the number of identical off-target hits, the number of off-target hits allowing for one or two mismatches, and the location on the genome. When hovering over the table, the currently active spacer is highlighted in the visualization, and vice versa.
6. Clicking on a table row adds the corresponding spacer to the download basket, a second click deselects the spacer again. The basket icon on the upper right of the screen displays the number of selected spacers.
7. Clicking the download basket icon takes the user to the download page. Here, a summary of the selected spacers is shown, and it is possible to download the selection as a comma-separated table that can be opened from a spreadsheet application or text editor.

3.2 Generation of a Random Size Deletion Library in Actinomycetes Using pCRISPR-Cas9 with Native NHEJ (See Note 2)

1. Digest the pCRISPR-Cas9 with Fast Digest NcoI and SnaBI or Eco105I restriction enzyme (Thermo Fisher Scientific, Waltham, US, Waltham, US). To prepare a stock of vector, a 100 μL reaction is used. Mix 5 μg (up to 10 μg) plasmid, 10 μL 10× Fast Digest Buffer, 5 μL of each of the Fast Digest restriction enzymes, and nuclease-free Milli-Q water to 100 μL. Incubate at 37 °C for 60 min. The digested plasmid is purified by GeneJET PCR Purification Kit (*see* **Note 3**). Then use NanoDrop 2000 to measure the concentration, and then Concentrator plus for concentrating the DNA solution when needed. The NcoI and SnaBI double-digested pCRISPR-Cas9 backbone solution can be stored in small aliquots at −20 °C for up to 6 months for multiple usages.
2. Identify spacers for functional sgRNA cassettes using CRISPy-web tool as described in Subheading 3.1; for each gene of interest, pick two spacers with minimal off-target effects.
3. Design primers for functional sgRNA cassette amplification, the forward primer can be designed as sgRNA-F: 5′-CATGCCATGGN$_{20}$GTTTTAGAGCTAGAAATAGC-3′ (N_{20} represents the 20 nt spacer sequence); the reverse primer stays the same as sgRNA-R: 5′-ACGCCTACGTAAAAAAAGCACCGACTCGGTGCC-3′. The restriction enzyme sites are underlined.
4. PCR is used to amplify the functional sgRNA cassette from pCRISPR-Cas9. 50 μL PCR reaction is used. Mix 20 ng (up to 100 ng) plasmid DNA, 10 μL 5× HF Buffer, 1 μL 10 mM dNTP mix, 0.5 μM of designed primers, 1.5 μL DMSO, 1 U Phusion Hot Start II DNA Polymerase, and nuclease-free Milli-Q water to 50 μL on ice, flip the PCR tubes by fingers, spin down the mixture. The PCR conditions are 98 °C for 30 s; 35 cycles of 98 °C for 10 s; T_a (up to 72 °C, calculated by Thermo Fisher Scientific T_m calculator from both primers) for 30 s; 72 °C for 10 s (1 kb/15–30 s); and finally 72 °C for 10 min, afterward keep at 4 °C.
5. Analyze the PCR products using 2% agarose gel on 1× TAE running buffer, the positive PCR product is purified by GeneJET PCR Purification Kit. Then use NanoDrop 2000 to measure the concentration, and Concentrator plus for concentrating the DNA solution when needed.
6. The purified PCR products are double digested by Fast Digest NcoI and SnaBI restriction enzymes, with the same condition of the plasmid double digestion.
7. Mix 100 ng of the double-digested pCRISPR-Cas9 backbone from **step 1**, fivefold of double-digested functional sgRNA cassette PCR product (from **step 6**), 1 μL 10× T4 Buffer,

and 1 U T4 DNA ligase in a total of 10 μL reaction volume, then incubate at 25 °C for 60 min.

8. Transform 50 μL of One Shot® Mach1™ T1 Phage-Resistant competent *E. coli* cells with 10 μL the ligation mixture, heat shock at 42 °C for 70 s, recover the cells in 300 μL SOC medium at 37 °C, 200 rpm for 1 h. Plate 200 μL of the recovered cells on selective LB agar plates with 50 μg/mL apramycin, and incubate at 37 °C overnight (around 16 h).
9. On the next day, pick 3–5 colonies into 0.5 mL selective LB liquid medium with 50 μg/mL apramycin in 1.5 mL Eppendorf tubes, incubate at 37 °C, 200 rpm for 4 h. 1 μL of each culture is used as a template for colony PCR validation of the ligation in **step 7**.
10. The colony PCR is carried out in a 20 μL reaction using Taq-based DNA polymerase. Mix 1 μL of the culture from **step 8** (20 ng of non-digested pCRISPR-Cas9 as a negative control), 10 μL of the PCR Master Mix (2×), 0.5 μM of the primers (sgRNA check-F: 5′-AATTGTACGCGGTCGATCTT-3′ and sgRNA check-R: 5′-TACGTAAAAAAAGCACCGAC-3′), and nuclease-free Milli-Q water to 20 μL on ice, flip the PCR tubes by fingers, spin down the mixture. Colony PCR conditions are 94 °C for 4 min; 35 cycles of 94 °C for 30 s; 50 °C for 30 s; 72 °C for 20 s (1 kb/1 min); and finally 72 °C for 10 min, then keep at 4 °C.
11. Analyze the PCR products using 4% agarose gel (*see* **Note 4**) on 1× TAE running buffer. Randomly pick two positive clones of each construct for 10 mL overnight culture using selective LB liquid medium with 50 μg/mL apramycin, at 37 °C.
12. On the next day, isolate the plasmids from the 10 mL culture of **step 10**, and confirm the results by Sanger sequencing using primer sgRNA check-F.
13. Transform ET12567/pUZ8002 competent *E. coli* cell with 100 ng of the validated plasmid using the same protocol as in **step 8**. Plate 100 μL of the recovered cells on selective LB agar plates with 50 μg/mL apramycin, 50 μg/mL kanamycin, and 25 μg/mL chloramphenicol, and incubate at 37 °C for around 24 h.
14. Randomly pick one clone (known as the donor strain for conjugation) from the selective LB plate of **step 12**, inoculate it into 10 mL of the same selective LB liquid medium, incubate at 37 °C overnight (around 24 h).
15. Wash the above culture twice with 10 mL LB liquid medium without antibiotics supplementation and then suspend the cell pellet with 1 mL (1/10 volume of the culture) LB liquid medium.

16. Mix 100 μL ET12567/pUZ8002 culture from **step 15** with 50 μL *S. coelicolor* A3(2) spores (*see* **Note 5**), and plate the mixture onto Cullum agar plates, inoculate the plates at 30 °C overnight (around 16 h).
17. On the next day, overlay the conjugation plates with 1 mL of sterilized Milli-Q water containing 1 mg nalidixic acid and 1 mg apramycin.
18. Incubate the plates at 30 °C for 3–5 days to let the exconjugants grow.
19. Pick customized number of exconjugants (the library size) and re-streak them onto ISP2 plates with 1 μg/mL thiostrepton, 50 μg/mL apramycin, and 50 μg/mL nalidixic acid for 5–7 days (*see* **Note 6**).
20. Inoculate the **step 19** strain into 20 mL non-antibiotic ISP2 liquid medium and incubate at 30 °C, 180 rpm for 3–5 days.
21. Isolate genomic DNA of the strains from **step 20** using Blood & Cell Culture DNA Kit.
22. The isolated genomic DNA can be used to analyze the random size deletion library.

3.3 Highly Efficient Gene Loss-of-Function Studies in Actinomycetes Using pCRISPR-Cas9-ScaligD (See Note 7)

Steps of **1**–**18** are identical to Subheading 3.2, except the backbone plasmid is pCRISPR-Cas9-ScaligD instead of pCRISPR-Cas9. The protocol diverges at **step 19**, when exconjugants can be seen from the conjugation plates.

19. Pick 10–20 exconjugants and re-streak them onto ISP2 plates with 1 μg/mL thiostrepton, 50 μg/mL apramycin, and 50 μg/mL nalidixic acid for 5–7 days.
20. Scratch some mycelia of the clones from **step 19** using a sterile toothpick into 10 μL pure DMSO (Sigma-Aldrich, St. Louis, US) in PCR tubes. Shake tubes vigorously for 10 min at 100 °C in a shaking heating block, vortex vigorously for another 1 min at room temperature, spin down the pellet at top speed for 10 s. Then 2 μL of the supernatant is used as a PCR template in a 50 μL reaction in **step 22**.
21. Design primers for amplifying an approximately 500 bp fragment, around the expected DSB site of mutation.
22. Mix 2 μL supernatant from **step 20**, 10 μL 5× GC Buffer, 1 μL 10 mM dNTP mix, 0.5 μM of designed primers from **step 21**, 1 U Phusion Hot Start II DNA Polymerase and nuclease-free Milli-Q water to 50 μL on ice, flip the PCR tubes by fingers, spin down the mixture. The PCR conditions are 98 °C for 30 s; 35 cycles of 98 °C for 10 s; T_a (up to 72 °C, is calculated by Thermo Fisher Scientific T_m calculator from both the primers) for 30 s; 72 °C for 10 s (1 kb/15–30 s); and finally 72 °C for 10 min, 4 °C forever.

23. Subclone the PCR products of **step 22** into pJET1.2/blunt vector from CloneJET PCR Cloning Kit, use pJET1.2 Forward Sequencing Primer from the kit for Sanger sequencing.
24. The desired mutations from **step 23** can be used for gene loss-of-function study.

3.4 Generation of In-Frame Gene/Gene Cluster Deletions or Replacements in Actinomycetes Using pCRISPR-Cas9 with Homologous Recombination Templates

As the homologous recombination is widely used for gene deletion and replacement in many organisms including *Streptomyces* [53]. We provide an editing template within the same plasmid, pCRISPR-Cas9 for HDR of the DSB caused by Cas9, to achieve scar-less genome editing.

Almost all the steps are the same as Subheading 3.2 (**steps 1–11** are exactly the same). The main different steps of this protocol are homologous recombination templates design and cloning.

12. Digest the pCRISPR-Cas9 that carries the designed spacer (the plasmid from **step 11** of Subheading 3.3) with Fast Digest StuI or Eco147I restriction enzyme. A 100 μL reaction volume is used. Mix 5 μg (up to 10 μg) plasmid, 10 μL 10× Fast Digest Buffer, and 5 μL of Fast Digest StuI restriction enzyme. Then incubate at 37 °C for 30 min. The linearized plasmid is purified by GeneJET PCR Purification Kit. Then use NanoDrop 2000 to measure the concentration and Concentrator plus for concentrating the DNA solution when needed.
13. Design primers for amplifying ~1 kb of both 5′ and 3′ fragments of the target gene (gene cluster), with 20 nt overhang at the end of both the fragments for later Gibson Assembly.
14. PCR amplify around 1 kb homologous recombination templates from genomic DNA of the WT strain. The 50 μL PCR reaction is used. Mix 150 ng (up to 500 ng) genomic DNA, 10 μL 5× GC Buffer, 1 μL 10 mM dNTP mix, 0.5 μM of designed primers, 1.5 μL DMSO, 1 U Phusion Hot Start II DNA Polymerase, and nuclease-free Milli-Q water to 50 μL on ice, flip the PCR tubes by fingers, spin down the mixture. The PCR conditions are 98 °C for 30 s; 35 cycles of 98 °C for 10 s; T_a (up to 72 °C, is calculated by Thermo Fisher Scientific T_m calculator from both the primers) for 30 s; 72 °C for 30 s (1 kb/15–30 s); and finally 72 °C for 10 min, then keep at 4 °C.
15. A 3-fragment Gibson assembly of 10 μL reaction volume is used to assemble the two homologous recombination templates into StuI site of pCRISPR-Cas9 with designed spacer. Mix 100 ng linearized plasmid (backbone), three- to fivefold of each of the two ~1 kb homologous recombination templates, 5 μL 2× Gibson Master Mix, and nuclease-free Milli-Q water to 10 μL in a PCR tube on ice. Flip the tube, spin down the mixture, incubate at 50 °C in a PCR block for 60 min.

16. Transform 50 μL of One Shot® Mach1™ T1 Phage-Resistant competent *E. coli* cells with 10 μL of each assembly reaction. Apply heat shock at 42 °C for 70 s, recover the cells in 300 μL SOC medium at 37 °C for 1 h. Plate 200 μL of the recovered cells on selective LB agar plates with 50 μg/mL apramycin, and incubate at 37 °C overnight (around 16 h).
17. On the next day, pick 3–5 colonies into 0.5 mL selective LB liquid medium with 50 μg/mL apramycin in 1.5 mL Eppendorf tubes, incubate at 37 °C, 200 rpm for 4 h for colony PCR using the designed primers which can cross both homologous recombination templates to validate the assembly in **step 15**.
18. The colony PCR is carried out in a 20 μL reaction, the same protocol as **step 9** of Subheading 3.3. Check the PCR products using 1% agarose gel on 1× TAE running buffer. Randomly pick two positive clones of each construct for 10 mL overnight culture. The following steps are the same as **steps 11–18**.
19. Randomly pick 3–5 exconjugants and restreak them onto ISP2 plates with 1 μg/mL thiostrepton, 50 μg/mL apramycin, and 50 μg/mL nalidixic acid for 5–7 days.
20. Inoculate the **step 19** strains into the 20 mL non-antibiotic ISP2 liquid medium and incubate at 40 °C, 180 rpm for 7 days to eliminate the CRISPR plasmid (*see* **Note 8**).
21. A proper (can be 1000- to 10,000-fold, dependents on the culture density) diluted fraction of **step 20** cultures is plated on non-antibiotic ISP2 plates to isolate single colonies.
22. Randomly pick 10–20 colonies of each strain from **step 21** and replica plate on ISP2 agar with and without 50 μg/mL apramycin. The clones with restored apramycin sensitivity have successfully eliminated the temperature-sensitive pCRISPR-Cas9 plasmid with the homologous recombination templates.
23. Scratch some mycelia of the plasmid-free clones from **step 22** using a sterile toothpick into 10 μL pure DMSO in PCR tubes. Let the tubes shake vigorously for 10 min at 100 °C in a shaking heating block, vortex vigorously for 1 min at room temperature, spin down the pellet at top speed for 10 s, 2 μL of the supernatant is used as the PCR template in a 50 μL reaction in **step 24** (*see* **Note 9**).
24. Mix 2 μL supernatant from **step 23**, 10 μL 5× GC Buffer, 1 μL 10 mM dNTP mix, 0.5 μM of designed primers from **step 17**, 1 U Phusion Hot Start II DNA Polymerase, and nuclease-free Milli-Q water to 50 μL on ice, flip the PCR tubes with fingers, spin down the mixture. The PCR conditions are 98 °C for 30 s; 35 cycles of 98 °C for 10 s; T_a (annealing temperature, up to 72 °C, was calculated by Thermo Fisher Scientific T_m calculator from the primers) for 30 s; 72 °C for *X* seconds (depends on

the amplicon size) (1 kb/15–30 s); and finally 72 °C for 10 min, keep on 4 °C.

25. Subclone the PCR products of **step 24** into pJET1.2/blunt vector from CloneJET PCR Cloning Kit, and use pJET1.2 Forward Sequencing Primer from the kit for Sanger sequencing.
26. The desired in-frame deletions can be identified from **step 25** (*see* **Note 10**).

3.5 pCRISPR-dCas9 for Gene Knockdown in Actinomycetes

For bacterial CRISPRi application, there are two locations of the DNA that can be targeted by dCas9:sgRNA complex to suppress the transcription; the gene coding region, and the region upstream of the start codon, which often includes the promoter. However, caused by a yet unknown mechanism, for sgRNAs targeting the coding region, only those sgRNAs that bind to the non-template DNA strand have the suppression effect, while no such strand bias is observed when the sgRNAs are targeting on the 5′-UTR [36, 38]. A simple illustration is shown in Fig. 3a, while a detailed example is in Fig. 3b [36].

The steps are the same as **steps 1–19** of Subheading 3.2. Except the spacer from the coding region of the sgRNAs needs to target on the non-template DNA strand.

20. The strains from **step 19** can be used for gene knock-down validation, either by measuring the final product or the mRNA level.

3.6 Introduction of USER Cassette (See Notes 11 and 12) into pCRISPR Serials Vector for the Facilitation of the CRISPR Vector Construction

The vector pGM1190-Cas9 is used to exemplify the de novo construction of a USER-compatible vector. In general, the USER vector contains a USER cassette composed of the *ccdB* gene-chloramphenicol resistance marker flanked by a restriction and a nicking enzyme. For the CRISPR-Cas9 system, thiostrepton and apramycin resistance markers are included in the vector for positive selection in *actinomyces* hosts. Two USER-compatible vectors are currently available in our lab: the pCRISPR-USER-Cas9 for gene knockout (in), and the pCRISPR-USER-dCas9 for gene knockdown.

1. The *ccdB*-chloramphenicol resistance marker cassette is amplified from the Gateway® Vector of the "Gateway® Vector Conversion System with One Shot® ccdB Survival Cells"-kit. For the PCR, mix 20 μL 5× Phusion® HF Buffer, 2 μL 10 mM dNTP mix, 1 μM of each primer (5′-AAAA**CGCCGGCG**GAATGCGTGCGATCGCAG-3′ and 5′-AAAA**GGGCCC**GAATGCACGCGATCGCTG-3′) (*see* **Note 13**), 2 U Phusion® HF DNA Polymerase, and nuclease-free Milli-Q water to 100 μL on ice, flip the PCR tubes by fingers, spin down the mixture. The PCR conditions are 98 °C for 30 s; 35 cycles of

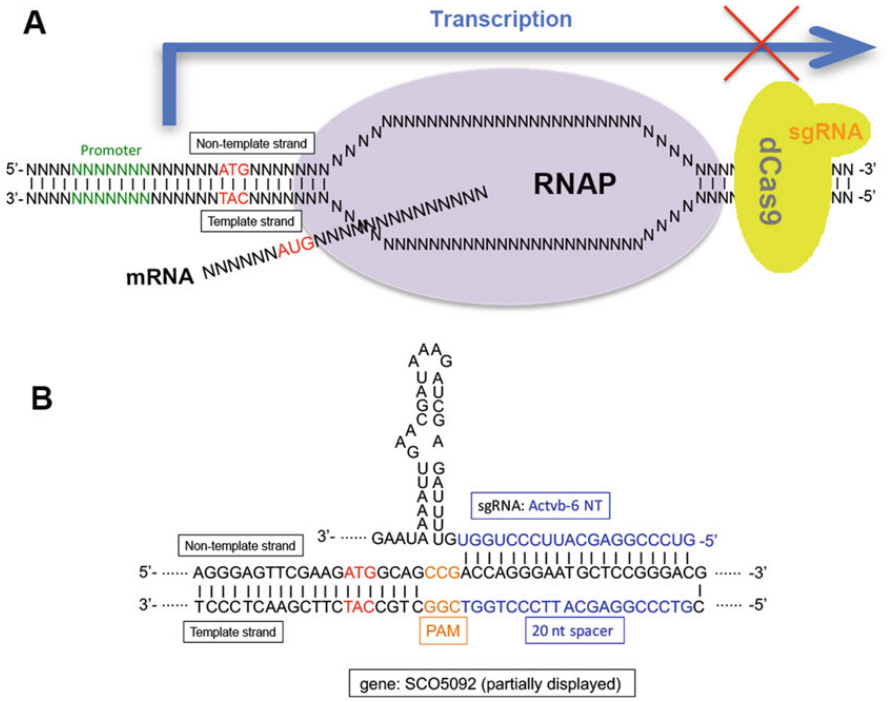


Fig. 3 Prokaryotic CRISPRi working model. (**a**) An overview of the prokaryotic CRISPRi working model. (**b**) Specific example of CRISPRi, the sgRNA is targeting on (binding to) the nontemplate DNA strand of the coding region of the SCO5092 gene in *S. coelicolor* [36]

98 °C for 10 s; 71 °C for 30 s; 72 °C for 1 min; and finally 72 °C for 10 min, keep on 4 °C.

2. Analyze the PCR product on 1% agarose gel with 1× TAE running buffer and purify the fragment using a gel cleanup kit.
3. Digest vector pGM1190-Cas9 and the PCR-amplified USER cassette with MreI and Bse120I. To ensure sufficient yields, digest up to 10 μg of the destination vector with 5 U of each enzyme in a total volume of 100 μL. For the PCR-amplified USER cassette, a 50 μL reaction with up to 3 μg DNA and 2 U of each enzyme normally is sufficient. The digestions are carried out at 37 °C. It is recommended to run the digestions overnight.
4. Analyze the digestions on 1% agarose gel with 1× TAE running buffer and purify the fragments with a gel cleanup kit.
5. For the ligation, mix digested vector and USER cassette in the ratio 1:3 (*see* **Note 14**). Mix 2 μL T4 DNA Ligase Reaction

Buffer, 1 U T4 DNA Ligase, and nuclease-free Milli-Q water in a total volume of 20 μL. The reaction can be carried out either at 25 °C for 1 h or 16 °C overnight and should be terminated by heating at 65 °C for 10 min.

6. Use 10 μL of the ligation mix for transformation of 50 μL One Shot® ccdB Survival™ 2 T1^{R} chemically competent cells. To account for variations in transformation efficiency, plate both 100 and 250 μL on pre-warmed selective LB plates with 50 μg/mL apramycin and 25 μg/mL chloramphenicol.
7. Verify clones by colony PCR and Sanger sequencing with primers covering the entire USER cassette (*see* **Note 15**). For the colony PCR, use 2 μL 10× DreamTaq buffer, 0.4 μL 10 mM dNTP mix, 0.1 μM of each primer (*see* **Note 15**), 2 μL colony (dissolved in 20 μL nuclease-free Milli-Q water), 0.2 U DreamTaq DNA Polymerase, and nuclease-free Milli-Q water to 20 μL on ice, flip the PCR tubes by fingers, spin down the mixture. PCR conditions are 95 °C for 3 min; 40 cycles of 95 °C for 30 s; 64 °C for 30 s; 72 °C for 2 min; and finally 72 °C for 5 min, keep on 4 °C. Analyze colony PCRs on 1% agarose with 1× TAE running buffer. Successful integration of the USER cassette should result in a PCR product of ~1.9 kb. Colonies with correct size of PCR product are subjected to Sanger sequencing. The final vector is named pCRISPR-USER-Cas9.
8. Linearize between 10 and 15 μg USER-compatible pCRISPR-USER-Cas9 with 20 U and 10 U Nb.BsmI in a total volume of 50 μL. Add 5 μL of the NEB3.1 or CutSmart® buffer and run the digestion at 37 °C for 2 h, followed by the inactivation of the enzymes at 65 °C for 1 h.
9. Run the 50 μL-reaction directly on 1% agarose gel with 1× TAE running buffer until a clear separation of the linearized vector and the *ccdB* + chloramphenicol resistance marker cassette (size of 1.7 kb).
10. Purify the vector with a gel cleanup kit (*see* **Note 16**).
11. GOIs are amplified using USER-compatible primers specific for the AsiSI/Nb.BsmI USER cassette. 7 to 12 nucleotides USER overhangs is sufficient for successful assembly [41]:

    5′-CGTGCGAU-[GOI_1]-3′.
    5′-CACGCGAU-[GOI_1]-3′.

    Use 20 ng gBlock DNA, and up to 200 ng genomic DNA as PCR template. For the PCR, mix 10 μL 5× Phusion GC Buffer with 1 μL 10 mM dNTP mix, 0.1 μM of each oligonucleotide, 1 U PfuX7 (*see* **Note 17**) or 1 U Phusion U Hot Start DNA Polymerase, and nuclease-free Milli-Q water to 50 μL on ice, flip the PCR tubes by fingers, spin down the mixture. The

PCR conditions are 98 °C for 3 min; 40 cycles of 98 °C for 10 s; T_a (up to 72 °C, is calculated by Thermo Fisher Scientific T_m calculator from both the primers) for 50 s; 72 °C for 10 s (1 kb/15–30 s); and finally 72 °C for 10 min, keep on 4 °C.

12. Analyze the PCR products on 1% agarose gel with 1× TAE running buffer and purify the right using a gel cleanup kit (*see* **Note 18**).
13. Mix the linearized USER-compatible vector with PCR fragments in the ratio 1:3 or 1:6. For PCR fragments <1000 bp use ratio 1:6 and for PCR fragments >1000 bp use ratio 1:3. Mix 1 μL USER™ enzyme mix and 0.5 μL 10× CutSmart® buffer with nuclease-free Milli-Q water in a total volume of 10 μL.
14. The USER excision is carried out at 37 °C for 15 min followed by 15 min at T_m of the USER overhang (if using AMUSER for USER overhang prediction software tools, then the T_m is 26 °C). Keep the mix at 10 °C for another 10 min to allow the DNA fragments be assembled into the vector.
15. For transformation, mix all 10 μL USER reaction with 90 μL NEB5-α competent *E. coli* cells and incubate the mixture on ice for 20 min before performing the heat-shock at 42 °C for 45 s. Incubate the transformation mixture on ice for 2 min followed by the addition of 250 μL pre-warmed (37 °C) SOC medium for recovering at 37 °C, 250 rpm for 1 h. Plate 100 and 250 μL (*see* **Note 19**) of the recovered cells on pre-warmed (37 °C) selective LB plates with appropriate antibiotics (in our case 50 μg/mL apramycin) and incubate at 37 °C overnight (about 16 h).
16. Perform colony PCR as described in **step 7**.

4 Notes

1. Before using the CRISPR-Cas9 toolkit in the strain of interest, please check that the pGM1190-based plasmid can replicate and that the strain codes for a *tipA* homologue that is required to activate the *tipA*-promoter.
2. This feature can only be used in the host with a defective NHEJ. The strain of interest needs to be checked first if there is a defective NHEJ and which component(s) are missing. For example, *S. coelicolor* A3(2) does not contain a gene encoding the ligase LigD.
3. The NcoI and SnaBI double digestion of pCRISPR-Cas9 generates two fragments, ~11 kb and 90 bp, because 90 bp is too small to be caught by the GeneJET PCR Purification Kit, we normally directly use GeneJET PCR Purification Kit to purify

the ~11 kb backbone, which will give you a higher yield. Of course, gel purification can be used to recycle the ~11 kb backbone.

4. Because the differences of successful ligation and self-ligation or the non-digested pCRISPR-Cas9 plasmid PCR are only 20 bp, which needs a high concentration of agarose gel to distinguish.
5. The amounts of ET strain and spore are highly case dependent, in this case, for *S. coelocolor A3(2)*, 10^8 ET strains and 10^7–10^8 spore per plate are sufficient to generate enough exconjugants, for other actinomyces strains, the conjugation condition needs to be customized.
6. Because the *tipA* promoter is leaky, we have observed that in some cases no induction with thiostrepton is required to provide sufficient Cas9 in the targeted cells.
7. This system, pCRISPR-Cas9-ScaligD, can only be functional in the host that is lack of DNA ligase component of NHEJ pathway.
8. For those actinomycetes that are high temperature sensitive: To eliminate the CRISPR plasmid, carry out **step 20** under 30 °C at least two times and then plate a proper diluted fraction of the culture on nonselective ISP2 plate to isolate single colonies. Check for apramycin sensitivity of these clones. If no candidate shows apramycin sensitivity, additional rounds of nonselective culture may be applied.
9. If colony PCR did not give you any positive bands, which may indicate that the colony PCR did not work, the genomic DNA needs to be isolated for PCR template.
10. If the two homologous recombination templates are outside the gene cluster, then this system can be used for the deletion of whole gene clusters/genomic regions. Alternatively, they can be designed to generate in frame deletions within single genes or used to introduce additional genes/DNA fragments at the targeted DNA region.
11. Instead, a simpler USER cassette can be designed in which the *ccdB* and chloramphenicol resistance marker cassette is omitted. For a detailed protocol on how to construct this type of cassette, *see* [54].
12. We use the *ccdB*-chloramphenicol resistance marker cassette as positive control for digestion but other systems might apply. The cassette can be amplified from the Gateway® Vector found in the Gateway® Vector Conversion System with One Shot® ccdB Survival Cells.
13. For de novo design of the USER cassette, design oligonucleotides with overhangs for *AsiSI* and *Nb.BsmI* as follows:

5′-AAAA-[MCS_enzyme1]-[Nb.BsmI]-[nt*]-[AsiSI]-3′.
5′-AAAA-[MCS_enzyme2]-[Nb.BsmI]-[nt*]-[AsiSI]-3′.

*The variable nucleotide (nt) ensures directional assembly during cloning.

14. To calculate the molar ratios, use a ligation calculator such as the one found at http://nebiocalculator.neb.com/#!/.
15. Examples of colony PCR and sequencing primers for the pCRISPR-Cas9/dCas9 constructs:

    USERseq_F: 5′-CGAGCGTCCGCCGGCG-3′.
    USERseq_internal_F: 5′-GCTAGTGTCATAGTCCTGAAAATCATCTG-3′.
    USERseq_internal_R: 5′-CTGGGTGAGTTTCACCAGTTTTGATTTAAAC-3′.
    USERseq_R: 5′-GCGTACCGCTTCGGGCCC-3′.

    Use USERseq_F and USERseq_R for colony PCRs and USERseq_internal_F and USERseq_internal_R for Sanger sequencing of the USER cassette.
16. To improve titers, elute DNA in smaller volumes and repeat the last elution step.
17. The PfuX7 DNA polymerase is prepared in-house by expressing the his-tagged protein followed by Ni-NTA chromatography. Hence, concentration of the enzyme can vary between batches and preliminary titrations are advised. Otherwise, the USER-compatible polymerase Phusion U is commercially available from Thermo Fisher Scientific.
18. We experience a higher frequency of correct transformants when gel purifying the PCRs prior to the USER assembly. However, for gBlock fragments it might prove sufficient with direct PCR purification using for example the GeneJET PCR Purification Kit.
19. Efficiency of the USER assembly depends on the number and sizes of the fragments for the assembly. If experiencing low efficiencies, consider plating all of the USER mixture on one plate.

Acknowledgments

This work was funded by grants from the Novo Nordisk Foundation to S.Y.L. and T.W. (NNF15OC0016226). The authors thank Günther Muth from the University of Tübingen for providing the pGM1190 plasmid.

References

1. Berdy J (2005) Bioactive microbial metabolites – a personal view. J Antibiot 58(1):1–26
2. Berdy J (2012) Thoughts and facts about antibiotics: where we are now and where we are heading. J Antibiot 65(8):385–395. doi:10.1038/Ja.2012.27
3. Hwang KS, Kim HU, Charusanti P et al (2014) Systems biology and biotechnology of *Streptomyces* species for the production of secondary metabolites. Biotechnol Adv 32(2):255–268. doi:10.1016/J.Biotechadv.2013.10.008
4. Blin K, Medema MH, Kazempour D et al (2013) antiSMASH 2.0-a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res 41(W1):W204–W212. doi:10.1093/nar/gkt449
5. Weber T, Blin K, Duddela S et al (2015) antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43(W1):W237–W243. doi:10.1093/nar/gkv437
6. Weber T, Kim HU (2016) The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. Synthetic Syst Biotechnol 1(2):69–79
7. Ziemert N, Alanjary M, Weber T (2016) The evolution of genome mining in microbes – a review. Nat Prod Rep 33(8):988–1005. doi:10.1039/c6np00025h
8. Weber T, Charusanti P, Musiol-Kroll EM et al (2015) Metabolic engineering of antibiotic factories: new tools for antibiotic production in actinomycetes. Trends Biotechnol 33(1):15–26. doi:10.1016/j.tibtech.2014.10.009
9. Deveau H, Garneau JE, Moineau S (2010) CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol 64:475–493. doi:10.1146/annurev.micro.112408.134123
10. Koonin EV, Makarova KS (2013) CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. RNA Biol 10(5):679–686. doi:10.4161/rna.24022
11. Barrangou R, Marraffini LA (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol Cell 54(2):234–244. doi:10.1016/j.molcel.2014.03.011
12. Abudayyeh OO, Gootenberg JS, Konermann S et al (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353(6299):aaf5573. doi:10.1126/science.aaf5573
13. Makarova KS, Wolf YI, Alkhnbashi OS et al (2015) An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol 13(11):722–736. doi:10.1038/nrmicro3569
14. Barrangou R, Fremaux C, Deveau H et al (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315(5819):1709–1712. doi:10.1126/science.1138140
15. Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet 45:273–297. doi:10.1146/annurev-genet-110410-132430
16. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327(5962):167–170. doi:10.1126/science.1179555
17. Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816–821. doi:10.1126/science.1225829
18. Deltcheva E, Chylinski K, Sharma CM et al (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471(7340):602–607. doi:10.1038/Nature09886
19. Nishimasu H, Ran FA, Hsu PD et al (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156(5):935–949. doi:10.1016/j.cell.2014.02.001
20. Sternberg SH, Redding S, Jinek M et al (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507(7490):62–67. doi:10.1038/Nature13011
21. Iliakis G, Wang H, Perrault AR et al (2004) Mechanisms of DNA double strand break repair and chromosome aberration formation. Cytogenet Genome Res 104(1–4):14–20. doi:10.1159/000077461
22. Kanaar R, Hoeijmakers JH, van Gent DC (1998) Molecular mechanisms of DNA double strand break repair. Trends Cell Biol 8(12):483–489
23. Bassett AR, Tibbit C, Ponting CP et al (2013) Highly efficient targeted mutagenesis of drosophila with the CRISPR/Cas9 system. Cell Rep 4(1):220–228. doi:10.1016/J.Celrep.2013.06.020
24. DiCarlo JE, Norville JE, Mali P et al (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res 41(7):4336–4343
25. Friedland AE, Tzur YB, Esvelt KM et al (2013) Heritable genome editing in *C. elegans* via a

CRISPR-Cas9 system. Nat Methods 10 (8):741–743. doi:10.1038/Nmeth.2532
26. Li DL, Qiu ZW, Shao YJ et al (2013) Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. Nat Biotechnol 31 (8):681–683. doi:10.1038/Nbt.2661
27. Mali P, Yang L, Esvelt KM et al (2013) RNA-guided human genome engineering via Cas9. Science 339(6121):823–826. doi:10.1126/science.1232033
28. Ronda C, Pedersen LE, Hansen HG et al (2014) Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. Biotechnol Bioeng 111(8):1604–1616. doi:10.1002/Bit.25233
29. Wang HY, Yang H, Shivalila CS et al (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153(4):910–918. doi:10.1016/J.Cell.2013.04.025
30. Xie KB, Yang YN (2013) RNA-guided genome editing in plants using a CRISPRCas system. Mol Plant 6(6):1975–1983. doi:10.1093/Mp/Sst119
31. Yang DS, Xu J, Zhu TQ et al (2014) Effective gene targeting in rabbits using RNA-guided Cas9 nucleases. J Mol Cell Biol 6(1):97–99. doi:10.1093/Jmcb/Mjt047
32. Lieber MR (1999) The biochemistry and biological significance of nonhomologous DNA end joining: an essential repair process in multicellular eukaryotes. Genes Cells 4(2):77–85
33. Deriano L, Roth DB (2013) Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Annu Rev Genet 47:433–455. doi:10.1146/annurev-genet-110711-155540
34. Aravind L, Koonin EV (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. Genome Res 11 (8):1365–1374. doi:10.1101/gr.181001
35. Bowater R, Doherty AJ (2006) Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. PLoS Genet 2 (2):e8. doi:10.1371/journal.pgen.0020008
36. Tong Y, Charusanti P, Zhang L et al (2015) CRISPR-Cas9 based engineering of actinomycetal genomes. ACS Synth Biol 4(9): 1020–1029. doi:10.1021/acssynbio.5b00038
37. Tong Y, Weber, T, Lee, SY (2015) CRISPR/Cas9-based system. EU Patent EP15160126.7, 25 Mar 2015
38. Qi LS, Larson MH, Gilbert LA et al (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152(5):1173–1183. doi:10.1016/J.Cell.2013.02.022
39. Czar MJ, Anderson JC, Bader JS et al (2009) Gene synthesis demystified. Trends Biotechnol 27(2):63–72. doi:10.1016/j.tibtech.2008.10.007
40. Nour-Eldin HH, Hansen BG, Norholm MHH et al (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res 34(18): e122. doi:10.1093/nar/gkl635
41. Bitinaite J, Rubino M, Varma KH et al (2007) USER (TM) friendly DNA engineering and cloning method by uracil excision. Nucleic Acids Res 35(6):1992–2002. doi:10.1093/nar/gkm041
42. Genee HJ, Bonde MT, Bagger FO et al (2015) Software-supported USER cloning strategies for site-directed mutagenesis and DNA assembly. ACS Synth Biol 4(3):342–349. doi:10.1021/sb500194z
43. Norholm MHH (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. BMC Biotechnol 10:21. doi:10.1186/1472-6750-10-21
44. Salomonsen BMUH, Halkier BA (2014) USER-derived cloning methods and their primer design. In: SLR V (ed) DNA cloning and assembly methods. Methods in molecular biology, vol 1116. Springer, New York
45. Cavaleiro AM, Kim SH, Seppala S et al (2015) Accurate DNA assembly and genome engineering with optimized uracil excision cloning. ACS Synth Biol 4(9):1042–1046. doi:10.1021/acssynbio.5b00113
46. Cobb RE, Wang Y, Zhao H (2014) High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. ACS Synth Biol. doi:10.1021/sb500351f
47. Huang H, Zheng GS, Jiang WH et al (2015) One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. Acta Biochim Biophys Sin (Shanghai) 47(4). doi:10.1093/abbs/gmv007
48. Zeng H, Wen SS, Xu W et al (2015) Highly efficient editing of the actinorhodin polyketide chain length factor gene in *Streptomyces coelicolor* M145 using CRISPR/Cas9-CodA(sm) combined system. Appl Microbiol Biotechnol 99(24):10575–10585. doi:10.1007/s00253-015-6931-4
49. Kieser T, Bibb M, Buttner M et al (2000) Practical *Streptomyces* genetics. Norwich, UK, The John Innes Foundation
50. Muth G, Nussbaumer B, Wohlleben W et al (1989) A vector system with temperature-

sensitive replication for gene disruption and mutational cloning in streptomycetes. Mol Gen Genet 219(3):341–348. doi:10.1007/Bf00259605

51. Xie SS, Shen B, Zhang CB et al (2014) sgRNAcas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. PLoS One 9(6). doi:10.1371/journal.pone.0100448
52. Blin K, Pedersen LE, Weber T et al (2016) CRISPy-web: an online resource to design sgRNAs for CRISPR applications. Synthetic Syst Biotechnol 1(2):4
53. Smithies O (2001) Forty years with homologous recombination. Nat Med 7(10): 1083–1086. doi:10.1038/nm1001-1083
54. Nour-Eldin HH, Geu-Flores F, Halkier BA (2010) USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories. In: AG F-N (ed) Plant secondary metabolism engineering. Methods in molecular biology, vol 643. Springer, New York

Manuscript III

An improved USER-CRISPR-Cas9 platform for high-throughput genome editing in streptomycetes

Helene Lunde Robertsen[1], Yaojun Tong[1], Ling Ding[1], Tilmann Weber[1], and Sang Yup Lee[1,2]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark.

[2]Department of Chemical and Biomolecular Engineering and BioInformatics Research Center, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea.

(Draft manuscript)

Abstract

Streptomycetes possess a great potential for production of clinically relevant secondary metabolites. However, in comparison to other industrial microorganisms, the molecular tools available for manipulating streptomycetes are still a limiting factor for large-scale metabolic engineering efforts.

The advent of the CRISPR-Cas9 genome engineering tool for actinomycetes has revitalised the field by now allowing for fast and efficient targeted editing of defined genes and gene clusters. The ease of handling and the general applicability of the tool favour its use in large-scale, high-throughput cloning settings, which allow for much greater output in terms of harnessing the full potential of promising isolates.

Here, we report the optimisation of a recently published cloning methodology termed the USER-CRISPR-Cas9 toolbox, suitable for fast and easy vector construction for targeted gene and gene cluster engineering in actinomycetes. Proof-of-concept was established for the optimised method based on the actinorhodin and undecylprodigiosin biosynthetic gene clusters in *Streptomyces coelicolor* A3(2). The improved cloning platform is compatible with an automated setting, which strengthens its applicability for future semi-automatic high-throughput genome engineering efforts in streptomycetes for improved production of valuable secondary metabolites.

Introduction

The gram-positive bacteria streptomycetes constitute the most well-described genus within the order of Actinomycetales[1]. Streptomycetes are primarily found in soil environments[1], however, other niches, such as freshwater and marine environments, have also been reported[2]. The bacteria are distinguished by several characteristics, including GC-rich genomes and a gifted secondary metabolite production profile, of which between 20 and 50 biosynthetic gene clusters (BGCs), spread on 0.8 – 3.0 Mb of the coding DNA, can be found in their genomes[3]. However, despite the genetic potential to synthesise multiple complex metabolites, only a fraction of the compounds are detected in routine fermentations. Instead, most of the BGCs, responsible for production of the secondary metabolites, remain silent under standard growth conditions. Although the reasons behind the silent nature of some clusters remain to be fully understood, the complex regulatory network used by these microorganisms to balance metabolism could play a role[4,5]. These regulatory circuits comprise global regulators, which can affect several pathways as a response to nutrient limitations or other environmental stresses, and local regulators, which act on specific pathways[6,7]. As more studies are undertaken to gain a greater understanding of the logic behind secondary metabolism in streptomycetes, more rational and targeted engineering solutions to close the gap between potential and actual production of novel compounds are becoming possible[8].

In addition to the increased understanding of secondary metabolite production and its regulation, the molecular toolbox used for genome manipulation of streptomycetes has also experienced great progress in recent years.

One of the more recent additions to the toolbox is the clustered regularly interspaced short palindrome repeats (CRISPR)-CRISPR-associated nuclease (Cas) system. CRISPR-Cas was first discovered as a part of the adaptive immune system in prokaryotes and archaea and displays similarities to the eukaryotic RNA interference (RNAi) mechanism[9]. The adaptive immunity is acquired based on the ability of the host to recognise incoming foreign DNA and store fragments, or protospacers, of this between short repetitive sequences in so-called CRISPR arrays in the host's genome[10]. The presence of a protospacer associated motif (PAM) upstream of the protospacer, found only in the foreign DNA and not in the CRISPR array, is believed to play a role in the acquisition stage as well[11]. Upon exposure to a virus or plasmid,

harbouring a protospacer sequence identical to a protospacer found in the CRISPR array, the foreign DNA is cleaved by the Cas proteins, which are guided by CRISPR RNA (crRNA), the mature form of precursor crRNA (pre-crRNA). The mechanism behind cleavage depends on the type of CRISPR system (type I – III) encoded in the host[10]. The type II CRISPR-Cas system is the most simple of the three types in that it requires only a limited set of components for interference[10,12]. This set includes the crRNA, *trans*-encoded small crRNA (tracrRNA), and the Cas9 protein. Here, the tracrRNA is responsible for the cleavage of pre-crRNA into crRNA, which then can form a complex with Cas9.

Recently, the type II CRISPR-Cas9 system has been adapted as a tool for genome engineering[13] and so far it has been implemented in a range of species, including, but not limited to, bacteria, yeast, mice, and human cell lines[14]. The CRISPR-Cas9 system has also been adapted for genome engineering of actinomycetes[15–17].

The two main components of the CRISPR-Cas9 editing system are the single guide RNA (sgRNA), which contains the 20 bp protospacer specific for a defined target sequence, and the endonuclease Cas9. Once the sgRNA binds to matching 20 bp sequence in the target DNA, Cas9 is recruited to cut the DNA, leaving a double strand break (DSB), which is repaired by the host by either non-homologous end joining (NHEJ) or homologous recombination (HR)[18]. The applicability of the system for genome engineering of actinomycetes has already been proven by successful deletions of genes as well as gene clusters[15–17], also in a multiplex manner[15], gene induction[19], and reversible repression of gene expression[17]. Here, the latter application relies on the catalytically inactive variant of Cas9 (dCas9), in which two point mutations (D10A and H840A) have been introduced in the RuvC and HNH domains of Cas9. As a result, the dCas9 can bind to the target DNA but lacks the endonuclease activity[13,20].

The potential of targeting any gene of interest (GOI) based on very few changes in the sgRNA makes the CRISPR-Cas9 system a highly valuable tools for generating libraries of mutants. However, despite the high efficiency of the system, the process of generating the constructs needed for the specific genome editing purposes can still be time-consuming. Hence, implementation of the plasmid construction in a semi-automatic high-throughput setting would improve the current experimental workflow and improve overall output. In this sense, the CRISPR-Cas9 system presents a highly

suitable system for automation in that it only requires co-expression of the sgRNA and the Cas9 protein, making the plasmid construction relatively straightforward.

So far, the exchange of 20 bp protospacers and HR templates, needed for the specific CRISPR-Cas9 constructs, has been achieved by different approaches, including Golden Gate assembly[15], Gibson assembly[16,17], and digestion and ligation-based cloning[15–17]. The uracil-specific excision reagent (USER)-cloning technique presents an excellent alternative to the current PCR-based cloning methods available for construction of the CRISPR-Cas9 constructs. The ligase-free USER-cloning technique is a highly efficient tool for directional, seamless assembly of multiple fragments in one step[21–23]. USER assembly requires only three components; fragments with PCR-generated 7 – 12 nucleotide (nt) overhangs flanked by uracil (dU) bases, a linearised vector with overhangs compatible to the fragment overhangs, and the USER™ enzyme, which is a mixture of a uracil DNA glycosylase (UDG) and the DNA glycosylase-lyase Endonuclease VIII[24,25]. Upon mixing, the USER™ enzyme recognises and cleaves dUs from the overhangs of the PCR-amplified fragments, leaving sticky ends compatible to those of the linearised vector (**Fig. 1**). The applicability of the USER assembly system has been greatly strengthened by the development of the commercial high-fidelity PfuTurbo® C_x Hotstart DNA polymerase[26] and the mutant *PfuX7* DNA polymerase[27], which are both capable of reading through the dU bases without loss of fidelity. The latest improvement to the system is the AMUSER 1.0 (http://www.cbs.dtu.dk/services/AMUSER/)[28], which is an *in silico* tool for designing oligonucleotides for the USER assembly. This web application and the simple features of the USER method makes it a suitable tool for construction of plasmids in a high-throughput manner in which a library of exchangeable control elements, such as promoters and terminators, can be introduced with genes of interest (GOIs) to generate new constructs ready for genome editing of a strain of interest.

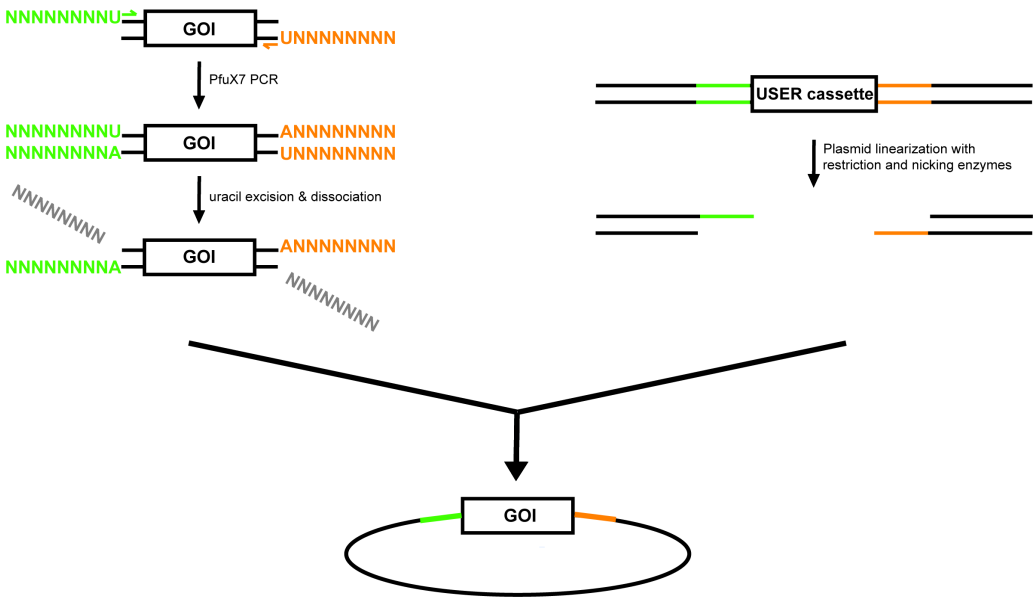


Figure 1 Schematic overview of USER assembly, which can be broken down into three main steps; First, the gene of interest (GOI) is PCR-amplified with oligonucleotides with 7 – 12 nucleotide overhangs flanked by a uracil (U) base. In parallel, the destination vector is linearisation with a restriction and a nicking enzyme, leaving overhangs complementary to those of the PCR-amplified fragments. In the third step, the GOI and linearised vector are mixed with the USER™ enzyme, yielding an assembled vector. GOI; gene of interest, USER; uracil-specific excision reagent.

Here we report the development of an optimised USER cloning platform for construction of CRISPR-Cas9 plasmids suitable for genome engineering of streptomycetes. Prior to implementation of the optimised system in an automated setting, the workflow was tested manually and evaluated based on the successful knock-down of actinorhodin (ACT) and undecylprodigiosin (RED) production in *Streptomyces coelicolor* A3(2) using the CRISPR-dCas9 construct. The next step will be to implement the system in a programmable automated workflow, which will allow for high-throughput plasmid construction.

With this optimised cloning platform we hope to strongly encourage future genome engineering efforts in streptomycetes in order to harness their full potential for production of novel and valuable secondary metabolites.

Materials and Methods

Bacterial strains and growth conditions

Plasmids and strains used in this study are listed in **Supplementary Table S1**.

All *Escherichia coli* transformations were carried out according to manufacturer's instructions and always included a 1 h recovery step in Super Optimal broth with Catabolite repression (S.O.C.) medium.

One Shot® *ccdB* Survival™ 2 T1R chemically competent cells (Thermo Fisher Scientific, Waltham, US) were used for propagation of pUSER plasmids harbouring the *ccdB* and chloramphenicol (CmR) gene cassette. Cells were propagated in Luria Broth (LB) medium supplemented with 50 μg/mL apramycin and 25 μg/mL chloramphenicol.

E. coli DH5-α cells were used for propagation of actinorhodin (ACT)- and undecylprodigiosin (RED)-CRISPR-dCas9 clones in LB supplemented with 50 μg/mL apramycin.

Prior to conjugation in *Streptomyces coelicolor* A3(2), plasmid DNA of ACT- and RED-CRISPR-dCas9 clones were introduced into *E. coli* ET12567(pUZ8002) chemically competent cells by calcium chloride transformation. ET12567 strains were propagated in LB medium supplemented with 50 μg/mL apramycin, 25 μg/mL kanamycin, and 25 μg/mL chloramphenicol for maximum 16 h.

S. coelicolor A3(2) was grown on modified SFM agar (2 % mannitol, 2 % full fat soy flour, 20mM $MgCl_2$, 20mM $CaCl_2$, tap water) for 5 – 7 days for sporulation. Intergeneric conjugation was carried out according to a standard protocol[29].

Construction of optimised USER-compatible pCRISPR vectors

Primers used in this study are listed in **Supplementary Table S2**. Furthermore, the sequence of gBlock® gene fragments, UC1 and UC2, and plasmid maps of pUSER1-dCas9, pUSER2-dCas9, pUSER2-Cas9, and pUSER2-Cas9/ScaLigD can be found in **Supplementary Notes S1** and **S2**.

The USER-compatible destination vectors pUSER1-Cas9, pUSER1-dCas9, and pUSER1-Cas9/ScaLigD were previously constructed according to[30]. To facilitate the use of the USER-compatible vectors in an automated setting, an alternative USER cassette (UC2) was introduced to replace the original cassette in the pUSER1

plasmids, yielding the three new plasmids pUSER2-Cas9, pUSER2-dCas9, and pUSER2-Cas9/ScaLigD. In the following, the plasmid construction will be exemplified with pUSER2-dCas9, since this plasmid was used for the proof-of-concept (PoC) studies reported here. To construct pUSER2-dCas9, UC2 was PCR-amplified using primers UP1/UP2 and introduced into the *MreI (Sse232I)/Bsp120I*-linearised pUSER1-dCas9. Ligation was carried out using T4 DNA ligase (New England Biolabs, NEB, Ipswich, US) following the manufacturer's instructions. Correct clones were verified by colony PCR using primers UP1/UP2 followed by Sanger sequencing with primers UP1 – UP4 to cover the entire USER cassette.

Construction of pUSER2-dCas9 plasmids for ACT and RED

The pUSER2-dCas9 plasmids for the knock-down of genes involved in biosynthesis of ACT and RED in *S. coelicolor* A3(2) were constructed as reported previously[30]. In brief, the 20 bp spacers specific for either SCO5087 (ActIorf1 in ACT) or SCO5878 (RedX in RED) were identified using CRISPy-web online tool[31] and corresponding primers for the USER assembly were predicted using AMUSER 1.0[28].

For the three ACT constructs ActIorf1-NT1, ActIorf1-NT2, and ActI-nontarget the sgRNA core sequence was PCR-amplified from UC1 using primers UP6/UP7/UP8 and UP11, of which the three forward primers contained the new 20 bp spacers to be introduced in front of the core sgRNA. For the RED constructs, UP9 – UP11 were used to PCR-amplify the sgRNAs containing the new 20 bp spacer specific for RedX (UP9) and a non-target spacer (UP10). The non-target 20 bp spacers were included for both ACT and RED to serve as negative controls. Destination vector pUSER2-dCas9 was linearised with restriction enzyme *AsiSI* and nicking enzyme *Nb.BsmI* (New England Biolabs, NEB, Ipswich, US). For the USER assembly, the sgRNA fragments and linearised pUSER2-dCas9 were mixed in the ratio 6:1 together with 1 unit USER™ enzyme (New England Biolabs, NEB, Ipswich, US). The mix was incubated at 37°C for 15 min, 25°C for 15 min, and finally 10°C for 10 min, upon which the mix was used for *E. coli* DH5α transformation. Correct clones were identified by colony PCR and verified by Sanger sequencing using primers UP12/UP13 (see **Supplementary Fig. S1**).

Detection of ACT and RED production in *S. coelicolor* A3(2) and pUSER2-dCas9 mutants

For induction of dCas9, exconjugants harbouring the ACT-CRISPR-dCas9 plasmids were streaked onto ISP2 agar containing 50 μg/mL apramycin, 25 μg/mL nalidixic acid, and 1 μg/mL thiostrepton, and incubated at 30°C for 7 days or until colonies became visible. ACT production of wild type strain *S. coelicolor* A3(2) and mutants A3(2)::ΔACT-orf1NT1/NT2 and A3(2)::ΔACT-nontarget were evaluated directly on the ISP2 plates containing 1 μg/mL thiostrepton for induction of the dCas9. Inhibition was confirmed by loss of the blue pigment characteristic for ACT production.

To evaluate RED production, wild type *S. coelicolor* A3(2), A3(2)::ΔRED-SCO5878, and A3(2)::ΔRED-nontarget strains were grown in 100 mL minimal medium[29] for 5 days with 1 μg/mL thiostrepton induction for the two mutant strains. RED production was determined using a modified protocol of Kang *et al*.[32]. 10 mL culture was spun down and cell pellet extracted with equal volume of absolute methanol. Extraction was carried out at 37°C for 1 h, shaking at 250 rpm, upon which cell debris was removed by centrifugation at 5000 *g* for 5 min. RED production was measured using an Orbitrap Fusion equipped with a Dionex Ultimate 3000 Ultra high performance liquid chromatography (UHPLC) pumping system (ThermoFisher Scientific, Waltham, MA, USA). Samples were held in the autosampler under 10°C during the analysis. 2 µL of each sample was injected to a C18 Acquity UPLC F5-3 HPLC column (2.1 × 100 mm, 1.8 µm,), at a flow rate of 0.4 mL/min, 30.0°C. Mobile phases A and B were 0.1 % formic acid in water and acetonitrile, respectively. Elution was done with a 30 min multistep system. After 5 % B for 1 min, a linear gradient started from 5 % B to 100% B in 21 min, which was held for another 5 min and followed by re-equilibration to 5 % B until 30 min. Profile data was collected in both positive and negative ion modes with a scan range of (m/z) = 100 – 1600. Further processing was carried out using Thermo Xcalibur™ 3.0 software.

Results

Construction of new CRISPR destination vectors allows for an optimised USER-cloning platform compatible with high-throughput settings

To take advantage of the simple features of USER assembly and CRISPR-(d)Cas9 genome engineering, three new plasmids were constructed. Prior to this study, the three USER-compatible vectors pUSER1-Cas9, pUSER1-dCas9, and pUSER1-Cas9/ScaLigD had already been constructed based on the backbones of pCRISPR-Cas9, pCRISPR-dCas9, and pCRISPR-Cas9-ScaLigD[17], respectively, and the protocol described by Tong, Robertsen *et al.*[30].

As the pUSER1 constructs required the exchange of the entire sgRNA (462 bp) for every new 20 bp protospacer chosen, the system had to be optimised to reduce costs and facilitate a standard assembly scheme. This was achieved by exchanging the original USER cassette (UC1), present in the pUSER1 constructs, with a new cassette (UC2), which was ordered as a gBlock® gene fragment (see **Supplementary Note S1**). In addition to the *ccdB* suicide gene and the chloramphenicol resistance marker (CmR), present in the original UC1, UC2 was expanded to include the sgRNA core flanked by the *ermE** promoter and the t_o terminator. The replacement of UC1 with UC2 in the original plasmids was facilitated through the *MreI* and *Bsp120I* sites flanking the cassettes. This resulted in the three new destination vectors pUSER2-Cas9, pUSER2-dCas9, and pUSER2-Cas9/ScaLigD (see **Supplementary Note S2**).

With the new system, exchanging the 20 bp spacer sequences is now facilitated through the USER primers (**Fig. 2**), in which the 20 bp are inserted directly after the 7 nt-dU overhangs. As a result, the organisation of the functional sgRNA is kept according to the original constructs[17] with the 20 bp spacer immediately downstream the constitutively expressed *ermE** promoter and upstream of the core sgRNA.

Using this new system, the price for the individual construct is drastically reduced. Instead of a 462 bp synthetic DNA fragment, the improved system only requires one generic reverse oligonucleotide (UP11) and one custom oligonucleotide containing the 20 bp spacer, the 7 nt-dU overhang, and 19 bp sequence complementary to the core sgRNA. pUSER2-Cas9 in addition can include the HR templates interspaced by the *ermE** promoter.

The optimised system, developed here, is suitable for high-throughput cloning in that the overall workflow requires limited operator interference and most of the individual

steps can be automated. Once supplied with the primers, templates, and linearised plasmid, the PCRs, USER assemblies, and transformations can be programmed on robotic equipment.

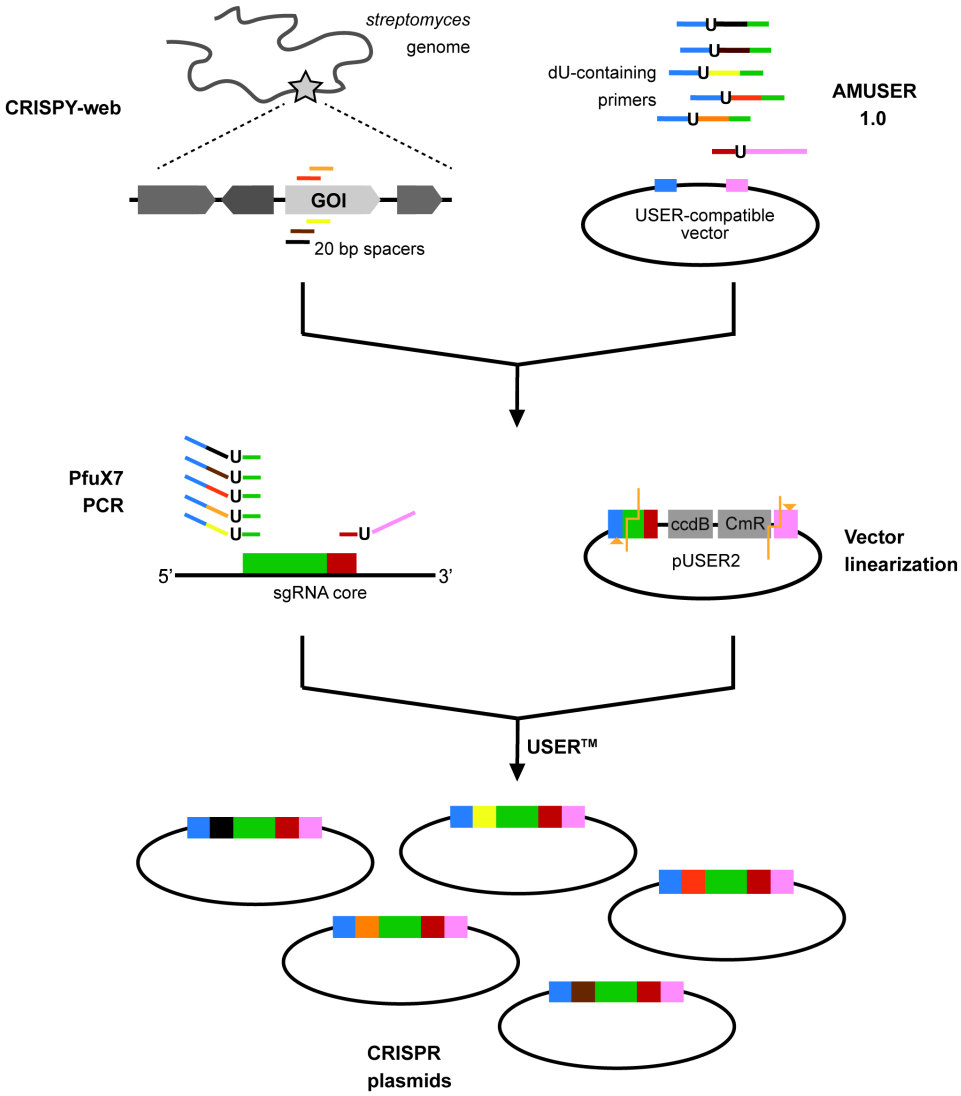


Figure 2 Schematic representation of the optimised USER-cloning method for construction of CRISPR knock-down (dCas) or knock-out (Cas9/Cas9-ScaLigD) plasmids. 20 bp recognition sequences (spacers) are predicted using CRISPy-web online tool. AMUSER 1.0 is used for prediction of the primers needed for the USER assembly of the sgRNA and the homologous recombination (HR) templates (the latter only for pUSER2-Cas9). Destinations vectors are linearised with a restriction and a nicking enzyme, leaving overhangs compatible with those of the USER-overhangs (upstream of uracil (U) base). Linearised vector and PCR-amplified uracil-containing fragments are assembled using the USER™ enzyme, which removes the dUs, resulting in generation of single-stranded overhangs compatible with the vector. CmR, chloramphenicol resistance marker; CRISPR, clustered regulatory interspaced short palindromic repeats; GOI, gene of interest; sgRNA, single-guide RNA; U, uracil; USER, uracil-specific excision reagent.

Repression of core genes involved in biosynthesis of ACT and RED in *S. coelicolor* A3(2) using engineered pUSER2-dCas9 constructs

To demonstrate the functionality of the new vector system, a proof-of-concept (PoC) was devised using pUSER2-dCas9 for gene knock-down of actinorhodin (ACT) production in *S. coelicolor* A3(2). The gene *actIorf1* (SCO5087), which encodes the ketosynthase (KS) and is a part of the minimal polyketide synthase (PKS) of ACT, was chosen as gene target. Two 20 bp spacers were tested; NT1, which had previously been used for successful gene knock-out[17], and NT2, predicted by CRISPy-web[31]. The results of the CRISPY-web prediction of protospacers for both SCO5087 and SCO5878 can be found in **Supplementary Fig. S2-I** and **S2-II**, respectively. Only spacers targeting the non-template strand of the genes *actorfI* and *redX* were chosen. Furthermore, spacers with few off-targets (defined as mis-matches in CRISPy-web) were preferred (see **Supplementary Fig. S2-I** and **S2-II**).

Using the optimised USER cloning system, the two plasmids pUSER2-dCas9::ΔACT-orf1NT1 and pUSER2-dCas9::ΔACT-orf1NT2 were constructed and used to generate the mutants A3(2)::ΔACT-orf1NT1 and A3(2)::ΔACT-orf1NT2, respectively. Two negative controls were included in the PoC as well; the non-target negative control A3(2)::ΔACT-nontarget, containing a 20 bp spacers targeting outside the ACT biosynthetic gene cluster (BGC), and the wild type strain *S. coelicolor* A3(2). The efficiency of the optimised USER assembly method was evident from the many positive clones in the colony PCR (see the gel in **Fig. 3**).

Disruption of the KS had previously been shown to lead to loss of the blue pigment in CRISPR-Cas9-generated mutants[17]. Therefore, for the initial PoC, the evaluation was based on visual examinations of the mutant strains and negative controls grown on ISP2 plates containing 1 μg/mL thiostrepton for the induction of dCas9 (**Fig. 3**). After seven days, visual examination of the streaked clones of A3(2)::ΔACT-orf1NT1 and A3(2)::ΔACT-orf1NT2 revealed a lack of ACT production as evident from the missing blue pigmentation (see **Supplementary Fig. S3**). Removal of the pUSER2-dCas9 plasmids by heat stressing resulted in restored blue pigmentation in the two ACT mutants (data not shown) and it was concluded that the loss of pigmentation was a result of the dCas9 gene repression of *actIorf1* involved in ACT biosynthesis in *S. coelicolor* A3(2).

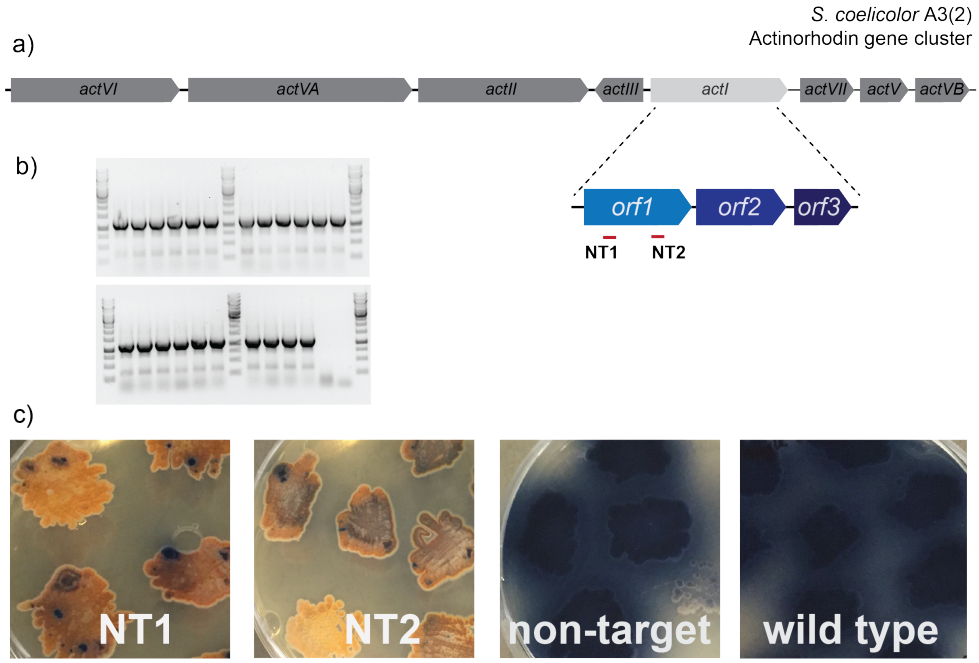
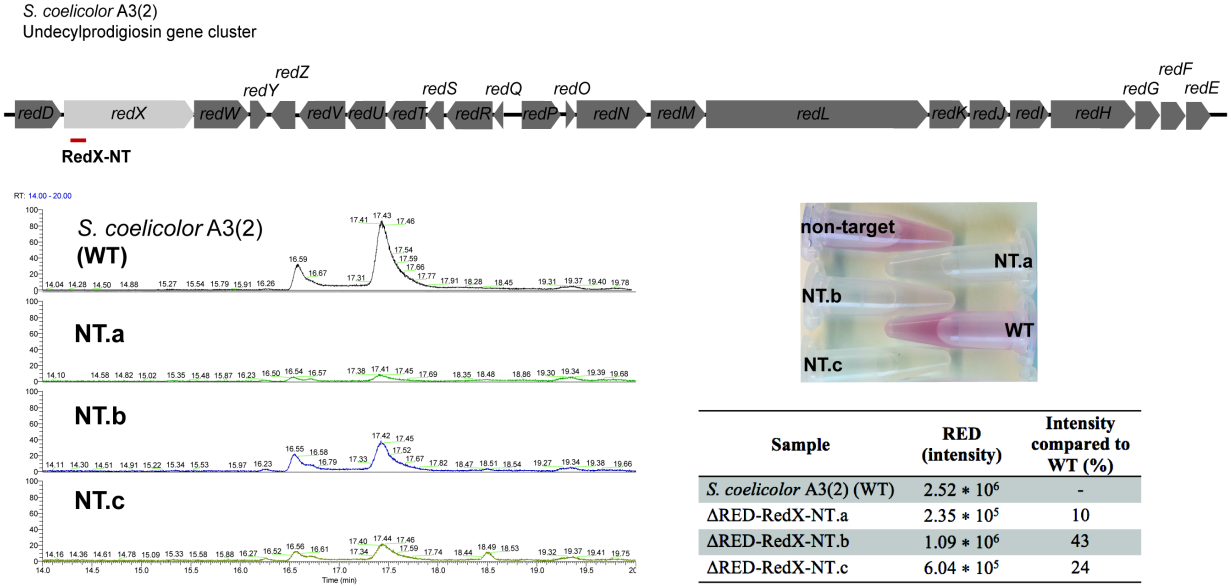


Figure 3 Proof-of-concept established using pUSER2-dCas9 for knock-down of actinorhodin (ACT) biosynthesis in *Streptomyces coelicolor* A3(2). **a)** The two 20 bp spacers NT1 and NT2, targeting the non-template strand of *actIorf1*, were designed and tested. **b)** pUSER2-dCas9 constructs were verified by colony PCR. The last two wells in the agarose gel are ddH_2O controls, otherwise all clones are positive. **c)** Pictures of the two mutants and the two negative controls grown on ISP2 with 1 µg/mL thiostrepton induction of dCas9 for seven days. The non-target 20 bp spacer was designed to guide the dCas9 outside the ACT cluster, hence, serving as a negative control. The wild type strain *S. coelicolor* A3(2) served as the second control. NT, non-template; orf, open reading frame.

Undecylprodigiosin (in the following referred to as RED) is another pigmented compound known to be produced by *S. coelicolor* A3(2). Again the pUSER2-dCas9 vector was chosen for knock-down of RED. This time, the 20 bp spacer was designed for *redX* (SCO5878), which encodes the two contiguous KSs, which have previously been proven essential for RED production in *S. coelicolor* A3(2)[33]. Similar to the previous approach, non-target and wild type negative controls were also included in the assay.

RED remains in the mycelium of *S. coelicolor* A3(2) and therefore only this was subjected to methanol extraction of the red-coloured pigment from the mutants and wild type[32]. The acidified (0.5 M HCl) methanol extracts were analysed by UHPLC – high-resolution mass spectrometry (HRMS). During the 30 min chromatographic run, RED appeared at 17.4 min (**Fig. 4**) with an m/z value of 393.565, corresponding to a molecular formula of $C_{25}H_{35}N_3O$. Based on this information, we could confirm successful knock-down of RED production in three biological replicates of A3(2)::ΔRED-RedX-NT (NT.a, NT.b, NT.c) when compared to the wild type sample.

Comparing the peak intensities (table in **Fig. 4**) revealed an average reduction in RED of 74 % in the mutant replicates compared to wild type. Furthermore, A3(2)::ΔRED-nontarget maintained RED production and it was concluded that the CRISPR/Cas9 strategy did not have polar effects on the RED BGC.



Sample	RED (intensity)	Intensity compared to WT (%)
S. coelicolor A3(2) (WT)	$2.52 * 10^6$	-
ΔRED-RedX-NT.a	$2.35 * 10^5$	10
ΔRED-RedX-NT.b	$1.09 * 10^6$	43
ΔRED-RedX-NT.c	$6.04 * 10^5$	24

Figure 4 Evaluation of undecylprodigiosin (RED) production in three biological replicates of A3(2)::ΔRED-RedX-NT (NT.a, NT.b, and NT.c). The 20 bp spacer was designed to target the non-template strand of *redX*, a core gene involved in RED biosynthesis in *Streptomyces coelicolor* A3(2). RED production inhibition was determined both by visual examination and HPLC-HRMS. The two negative controls include *S. coelicolor* A3(2) wild type (WT) and a non-target control, in which the 20 bp spacer was designed to guide the dCas9 outside of the RED cluster. Furthermore, RED production was evaluated based on recorded intensities and percentage (%) intensity compared to WT RED production (table). NT, non-template; WT, wild type.

Discussion

Genome engineering can be an efficient way of changing properties of streptomycetes and improving their production of secondary metabolites. In this context, the development of a CRISPR-Cas9 genome engineering tool for actinomycetes has revitalised our efforts. Here, we report the expansion of an existing CRISPR-Cas9 editing toolbox[30] through the development of an optimised USER-assembly system for construction of CRISPR-(d)Cas9 plasmids. The new workflow allows for faster and cheaper plasmid construction, and furthermore, is compatible with a high-throughput semi-automatic platform.

In this study, we used the properties of pUSER2-dCas9-catalysed gene repression for fast and easy evaluation of ACT and RED production in *S. coelicolor* A3(2) mutants. However, for actual gene and gene cluster deletions we also have available in our group the destination vectors pUSER2-Cas9 and pUSER2-Cas9/ScaLigD. The properties of these two vectors allow for construction of not only gene specific deletion constructs, but could also can aid in fast construction of genome-minimised *Streptomyces* hosts in a multiplexing manner[15]. The incentive to construct genome-minimised *Streptomyces* hosts includes the potential of such strains to become scaffolds for heterologous expression of otherwise silent gene clusters, originating from strains, that are not easily cultivated under standard laboratory conditions. Removing non-essential parts of the chromosome, most often dedicated to secondary metabolism, can relieve precursors for more desired pathways, and hence, serve as an important field of research when it comes to production of novel compounds or titer improvements[34].

Furthermore, the general applicability of the CRISPR-Cas9 system also allows for its use in engineering efforts to induce production of secondary metabolites in streptomycetes. In this case, both dCas9 and Cas9 can be used to target regulatory elements, such as repressors or promoters, found in silent BGCs in an attempt to induce production of potentially novel compounds. Recently, an example of taking advantage of this alternative approach was reported by Zhang *et al.*[19], who used the CRISPR-Cas9 system for knock-in of constitutive promoters in front of several BGCs in different *Streptomyces* hosts, triggering the production of new metabolites from otherwise silent BGCs.

Our improved pUSER2 vector toolbox is compatible with a semi-automatic high-throughput platform. In practise, the entire USER assembly can be automated in that it only requires the primers with USER-overhangs, templates for the *Pfu*X7 PCRs, and the linearised vectors. In this case, the PCR-amplified fragments can even be used directly for the USER assembly without any prior purification needed[35] (**Fig. 5**).
To further strengthen the applicability of the USER-CRISPR-Cas9 cloning platform, the inclusion of a chromogenic selection marker could aid in selection of correct clones. One such system could be the GusA reporter system, which has already been optimised for *Streptomyces* genome engineering[36]. The gene *gusA* encodes a β-glucoronidase, which can cleave the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-Gluc) into the blue precipitate 5,5'-dibromo-4,4'-dichloro-indigo. Including the *gusA* gene in the pUSER2 plasmids would allow for quick selection of *Streptomyces* clones harbouring the plasmid and further would reduce the time spent screening for clones having lost the plasmids after Cas9/dCas9 induction has occurred.

ACT and RED production can be readily monitored in *S. coelicolor* A3(2) due to the visible change in pigmentation, and, therefore, presents two highly applicable biological systems for PoC studies. However, since most secondary metabolites lack visual markers, the analytics required in screening for these desired or new compounds presents a time-limiting factor. In such cases, an untargeted metabolomics approach, which is the complete analysis of all metabolites detectable in a sample[37], might be employed for evaluation of the genome engineering strategy. This methodology allows for detection of not only expected compounds, but also potentially novel ones. Several databases exist for dereplication of natural products in complex sample mixtures, including the open-access Global Natural Products Social Molecular Networking (GNPS) knowledge base, which contains community-derived MS libraries of raw, processed, and identified tandem mass (MS/MS) spectrometry data[38], or alternatively, the Global Alignment for natural-products chemInformatiCs (GARLIC) and Generalized Retrobiosynthetic Assembly Prediction Engine (GRAPE) platform, which combines compound mining with automated compound characterisation[39]. Linking the output from the engineering efforts to a community-driven compound database would allow for constant re-evaluation of the sample

mixtures and could improve dereplication efforts on the long haul as the database is expected to undergo constant expansion.

In this study, the overall workflow of the USER-CRISPR-Cas9 system was tested and verified manually by genetic manipulation of the ACT and RED clusters in *S. coelicolor* A3(2). The next step now is implementation of the system in a robotic setting to evaluate the potential of the workflow for high-throughput plasmid construction. If successful, we believe that the optimised system can greatly improve our future genome engineering efforts. Implementation of the extensions and improvements discussed above could further improve the versatility of the system, although we believe that the current USER-CRISPR-Cas9 cloning system, reported in this study, already serves as a nice tool for future strain engineering efforts and ultimately could improve our ability to harness the potential of streptomycetes for secondary metabolite production.

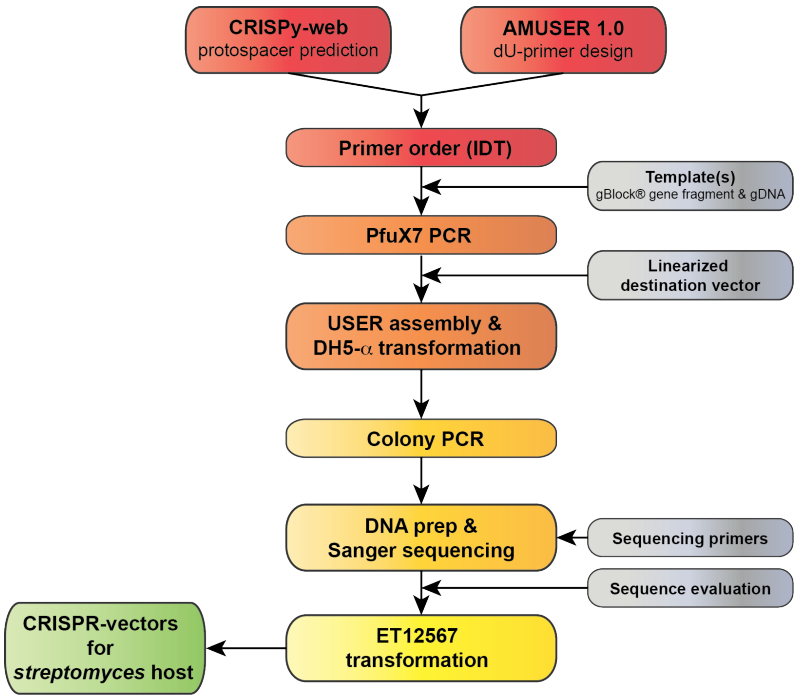


Figure 5 Proposed semi-automatic high-throughput workflow for the USER-assembly of pUSER2 vectors. Inputs, supplied externally by the user, are displayed on the right side of the workflow and output, in this case CRISPR-vectors ready for introduction into *streptomyces* hosts of choice, on the left side. The individual steps presented in the middle of the workflow should be automated to ensure a high-throughput output. AMUSER, automated DNA modifications with USER cloning. CRISPR, clustered regulatory interspaced short palindromic repeats; IDT, Integrated DNA Technologies; gDNA, genomic DNA; USER, uracil-specific excision reagent.

Acknowledgments

This work was funded by the Novo Nordisk Foundation.

References

1. Barka, E. A. *et al.* Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **80,** 1–43 (2016).

2. Valliappan, K., Sun, W. & Li, Z. Marine actinobacteria associated with marine organisms and their potentials in producing pharmaceutical natural products. *Appl. Microbiol. Biotechnol.* **98,** 7365–7377 (2014).

3. Baltz, R. H. Gifted microbes for genome mining and natural product discovery. *J. Ind. Microbiol. Biotechnol.* **44,** 573–588 (2017).

4. Bibb, M. J. Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* **8,** 208–215 (2005).

5. Baltz, R. H. Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes. *J. Ind. Microbiol. Biotechnol.* **43,** 343–370 (2016).

6. van Wezel, G. P. & McDowall, K. J. The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Nat. Prod. Rep.* **28,** 1311–1333 (2011).

7. Liu, G., Chater, K. F., Chandra, G., Niu, G. & Tan, H. Molecular Regulation of Antibiotic Biosynthesis in *Streptomyces*. *Microbiol. Mol. Biol. Rev.* **77,** 112–143 (2013).

8. Robertsen, H. L., Weber, T., Kim, H. U. & Lee, S. Y. Towards systems metabolic engineering of streptomycetes for secondary metabolites production. *Biotechnol. J.* 1700465 (2017). doi:10.1002/biot.201700465

9. Makarova, K. S., Grishin, N. V, Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1,** 7 (2006).

10. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annu. Rev. Genet.* **45,** 273–297 (2011).

11. Deveau, H., Garneau, J. E. & Moineau, S. CRISPR/Cas System and Its Role in Phage-Bacteria Interactions. *Annu. Rev. Microbiol.* **64,** 475–493 (2010).

12. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science (80-.).* **315,** 1709–1712 (2007).

13. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337,** 816–821 (2012).

14. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32,** 347–355 (2014).

15. Cobb, R. E., Wang, Y. & Zhao, H. High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **4,** 723–728 (2015).

16. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim. Biophys. Sin. (Shanghai).* **47,** 231–243 (2015).

17. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* **4,** 1020–1029 (2015).

18. Kanaar, R., Hoeijmakers, J. H. . & van Gent, D. C. Molecular mechanisms of DNA double-strand break repair. *Trends Cell Biol.* **8,** 483–489 (1998).

19. Zhang, M. M. *et al.* CRISPR–Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters. *Nat. Chem. Biol.* **13,** 607–609 (2017).

20. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).

21. Nisson, P. E., Rashtchian, A. & Watkins, P. C. Rapid and efficient cloning of Alu-PCR products using uracil DNA glycosylase. *Genome Res.* **1,** 120–123 (1991).

22. Bitinaite, J. *et al.* USER friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res.* **35,** 1992–2002 (2007).

23. Nour-Eldin, H. H., Geu-Flores, F. & Halkier, B. A. in 185–200 (2010). doi:10.1007/978-1-60761-723-5_13

24. Melamede, R. J., Hatahet, Z., Kow, Y. W., Ide, H. & Wallace, S. S. Isolation and Characterization of Endonuclease VIII from *Escherichia coli*. *Biochemistry* **33,** 1255–1264 (1994).

25. Jiang, D., Hatahet, Z., Melamede, R. J., Kow, Y. W. & Wallace, S. S. Characterization of *Escherichia coli* Endonuclease VIII. *J. Biol. Chem.* **272,** 32230–32239 (1997).

26. Nour-Eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K. & Halkier, B. A. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* **34,** e122 (2006).

27. Nørholm, M. H. *et al.* A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol.* **10,** 21 (2010).

28. Genee, H. J. *et al.* Software-Supported USER Cloning Strategies for Site-Directed Mutagenesis and DNA Assembly. *ACS Synth. Biol.* **4,** 342–349 (2015).

29. Kieser, T. Practical *Streptomyces* genetics. (2000).

30. Tong, Y., Robertsen, H. L., Blin, K., Weber, T. & Lee, S. Y. in *Methods in Molecular Biology* **1671,** 163–184 (2018).

31. Blin, K. CRISPy-web: An online resource to design sgRNAs for CRISPR applications. *Synth.*

Syst. Biotechnol. **1,** 118–121 (2016).

32. Kang, S. Actinorhodin and undecylprodigiosin production in wild-type and *relA* mutant strains of *Streptomyces coelicolor* A3(2) grown in continuous culture. *FEMS Microbiol. Lett.* **168,** 221–226 (1998).

33. Cerdeño, A. M., Bibb, M. J. & Challis, G. L. Analysis of the prodiginine biosynthesis gene cluster of *Streptomyces coelicolor* A3(2): new mechanisms for chain initiation and termination in modular multienzymes. *Chem. Biol.* **8,** 817–829 (2001).

34. Bilyk, O. & Luzhetskyy, A. Metabolic engineering of natural product biosynthesis in actinobacteria. *Curr. Opin. Biotechnol.* **42,** 98–107 (2016).

35. Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T. & Nørholm, M. H. H. Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth. Biol.* **4,** 1042–1046 (2015).

36. Myronovskyi, M., Welle, E., Fedorenko, V. & Luzhetskyy, A. Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. *Appl. Environ. Microbiol.* **77,** 5370–5383 (2011).

37. Frédérich, M., Pirotte, B., Fillet, M. & de Tullio, P. Metabolomics as a Challenging Approach for Medicinal Chemistry and Personalized Medicine. *J. Med. Chem.* **59,** 8649–8666 (2016).

38. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34,** 828–837 (2016).

39. Dejong, C. A. *et al.* Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat. Chem. Biol.* **12,** 1007–1014 (2016).

Supplementary Information

An improved USER-CRISPR-Cas9 platform for high-throughput genome editing in streptomycetes

Helene Lunde Robertsen[1], Yaojun Tong[1], Ling Ding[1], Tilmann Weber[1], and Sang Yup Lee[1,2]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark.

[2]Department of Chemical and Biomolecular Engineering and BioInformatics Research Center, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea.

Table of Contents

Table S1: Plasmids and strains used in the study

Source	Description/Genotype	Reference
	***E. coli* strains**	
One Shot® *ccdB* Survival™ 2 T1R chemically competent cells	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*Δ139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G *fhuA*::*IS2*	Thermo Fisher Scientific
E. coli DH5-α	*fhu*A2, Δ(*arg*F-*lac*Z)U169, *pho*A, *gln*V44, Φ80, Δ(*lac*Z)M15, *gyr*A96, *rec*A1, *rel*A1, *end*A1, *thi*-1, *hsd*R17	New England Biolabs
E. coli ET12567(pUZ8002)	*dam*13::Tn9(ChlR), *dcm*-6, *hsd*M, *hsd*R, *rec*F143, *zjj*-201::Tn10, *gal*K2, *gal*T22, *ara*14, *lac*Y1, *xyl*-5, *leu*B6, *thi*-1, *ton*A31, *rps*L136, *his*G4, *tsx*-78, *mtl*I, *gln*V44. pUZ8002(KanR)	[1]
	***Streptomyces* strains**	
Streptomyces coelicolor A3(2)	Wild type strain	[2]
A3(2)::ΔACT-orf1NT1	*S. coelicolor* A3(2) harbouring pUSER2-dCas9::ACT-orf1NT1. ApraR	This study
A3(2)::ΔACT-orf1NT2	*S. coelicolor* A3(2) harbouring pUSER2-dCas9::ACT-orf1NT2. ApraR	This study
A3(2)::ΔACT-nontarget	*S. coelicolor* A3(2) harbouring pUSER2-dCas9::ACT-nontarget. ApraR	This study
A3(2)::ΔRED-RedX-NT	*S. coelicolor* A3(2) harbouring pUSER2-dCas9::RED-RedX-NT. ApraR	This study
A3(2)::ΔRED-nontarget	*S. coelicolor* A3(2) harbouring pUSER2-dCas9::RED-nontarget. ApraR	This study
	Plasmids	
pUSER1-Cas9	USER-compatible destination vector, which is a derivative of pGM1190 with Cas9 under the control of TipA promoter. Contains USER cassette with *ccdB* suicide gene and chloramphenicol resistance marker. ApraR, CmR	[3]
pUSER1-dCas9	USER-compatible destination vector, which is a derivative of pGM1190 with dCas9 under the control of TipA promoter. Contains USER cassette with *ccdB* suicide gene and chloramphenicol resistance marker. ApraR, CmR	[3]
pUSER1-Cas9/ScaLigD	USER-compatible destination vector, which is a derivative of pGM1190 with Cas9 under the control of TipA promoter and ScaLigD under the control of *ermE** promoter. Contains USER cassette with *ccdB* suicide gene and chloramphenicol resistance marker. ApraR, CmR	[3]
pUSER2-Cas9	USER-compatible destination vector, which is a derivative of pUSER1 with Cas9 under the control of TipA promoter. Contains USER cassette with sgRNA, *ccdB* suicide gene, and chloramphenicol resistance marker. ApraR, CmR	This study
pUSER2-dCas9	USER-compatible destination vector, which is a derivative of pUSER1 with dCas9 under the control of TipA promoter. Contains USER cassette with	This study

	sgRNA, *ccdB* suicide gene, and chloramphenicol resistance marker. ApraR, CmR	
pUSER2-Cas9/ScaLigD	USER-compatible destination vector, which is a derivative of pUSER1 with Cas9 under the control of TipA promoter and ScaLigD under the control of ermE* promoter. Contains USER cassette with sgRNA, *ccdB* suicide gene, and chloramphenicol resistance marker. ApraR, CmR.	This study
pUSER2-dCas9::ACT-orf1NT1	pUSER2-dCas9 with USER cassette exchanged with sgRNA with 20 bp recognition sequence for non-template (NT) 1 for ActIorf1 (SCO5087) involved in actinorhodin biosynthesis. ApraR	This study
pUSER2-dCas9::ACT-orf1NT2	pUSER2-dCas9 with USER cassette exchanged with sgRNA with 20 bp recognition sequence non-template (NT) 2 for ActIorf1 (SCO5087) involved in actinorhodin biosynthesis. ApraR	This study
pUSER2-dCas9::ACT-nontarget	pUSER2-dCas9 with USER cassette exchanged with sgRNA with a 20 bp recognition sequence, which does not target (non target) genes involved in actinorhodin biosynthesis. ApraR	This study
pUSER2-dCas9::RED-RedX-NT	pUSER2-dCas9 with USER cassette exchanged with sgRNA with 20 bp recognition sequence for RedX (SCO5878) involved in undecylprodigiosin biosynthesis. ApraR	This study
pUSER2-dCas9::RED-nontarget	pUSER2-dCas9 with USER cassette exchanged with sgRNA with a 20 bp recognition sequence, which does not target (non target) genes involved in undecylprodigiosin biosynthesis. ApraR	This study

Table S2: Primers used in this study

Primer ID	Primer name	Sequence (5' - 3')	PCR program
UP1	UC2_MreI_FW	AAAA**CGCCGGCG**GCTAGCAAAGC GGTCGATCTTGACGGCTG	1. 98°C, 0:30 min 2. 98°C, 0:10 min 3. 70°C, 0:30 min 4. 72°C, 1:10 min 5. 72°C, 5:00 min (Steps 2-4 × 34)
UP2	UC2_Bsp120I_RV	AAAA**GGGCCC**GAATGCACGCGAT CGCTGCGCCG	
UP3	UC2_intseq_FW	GTCGTTCGGCTTCATCTGGATTTTC AG	1. 95°C, 3:00 min 2. 95°C, 0:30 min 3. 64°C, 0:30 min 4. 72°C, 1:00 min 5. 72°C, 5:00 min (Steps 2-4 × 40)
UP4	UC2_intseq_RV	GGCGTGTTACGGTGAAAACCTGG	
UP6	ActIorf1-NT1_FW	**CGTGCGAU**GTGGCTCGAAGGAGG CTCGAGTTTTAGAGCTAGAAATAG	1. 98°C, 0:30 min 2. 98°C, 0:10 min 3. 68°C, 0:30 min 4. 72°C, 0:15 min 5. 72°C, 5:00 min (Steps 2-4 × 34)
UP7	ActIorf1-NT2_FW	**CGTGCGAU**ATCGACTTGATCGAG CTGACGTTTTAGAGCTAGAAATAG	
UP8	ActI-nontarget_FW	**CGTGCGAU**GTCCGGGTCGAGCCC GCTGGGTTTTAGAGCTAGAAATAG	
UP9	RedX-NT1_FW	**CGTGCGAU**GTGGGTGTGTCCGTG GCTGTGTTTTAGAGCTAGAAATAG	
UP10	RedX-nontarget_FW	**CGTGCGAU**CTTCGGCAGTTCCAA CATCAGTTTTAGAGCTAGAAATAG	
UP11	sgRNA_RV	**CACGCGAU**CCTCAGGTGACCTCA GAACTCCATCTG	
UP12	USERseq_FW	GTTCGAGGTCATGTCCGAGGAG	1. 95°C, 3:00 min 2. 95°C, 0:30 min 3. 64°C, 0:30 min 4. 72°C, 1:00 min 5. 72°C, 5:00 min (Steps 2-4 × 40)
UP13	USERseq_RV	CCGTCTGACGCCCGATCACG	

Figure S1: Verification of pUSER2-dCas9 clones by colony PCR

Colony PCR was carried out to verify correct USER assembly of sgRNAs, amplified with primer containing new 20 bp recognition sequences placed in front of the core sgRNA, in pUSER2-dCas9. Correct clones (**1**) yield an amplified fragment of 1.1 kb whereas religated pUSER-dCas9 plasmid (**2**) yields a 2.5 kb fragment. The two scenarios are shown in the gel on the right in the figure below. GeneRuler 1 kb DNA Ladder (**L**) was used as marker in which the two intense bands correspond to 1 (lower) and 3 (upper) kb.

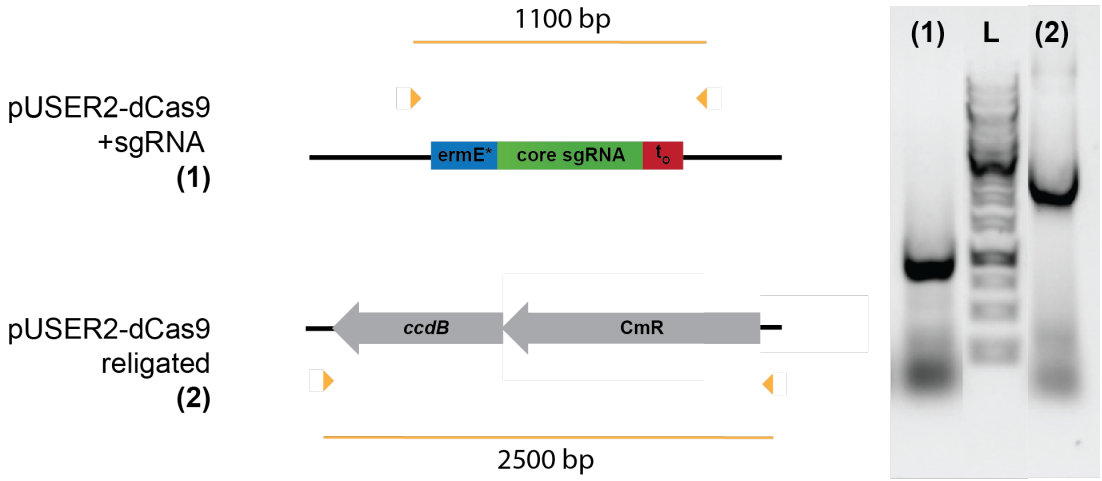


Figure S2-I – S2-II: CRISPy-web prediction of spacers for SCO5087 (ACT) and SCO5878 (RED) in *Streptomyces coelicolor* A3(2)

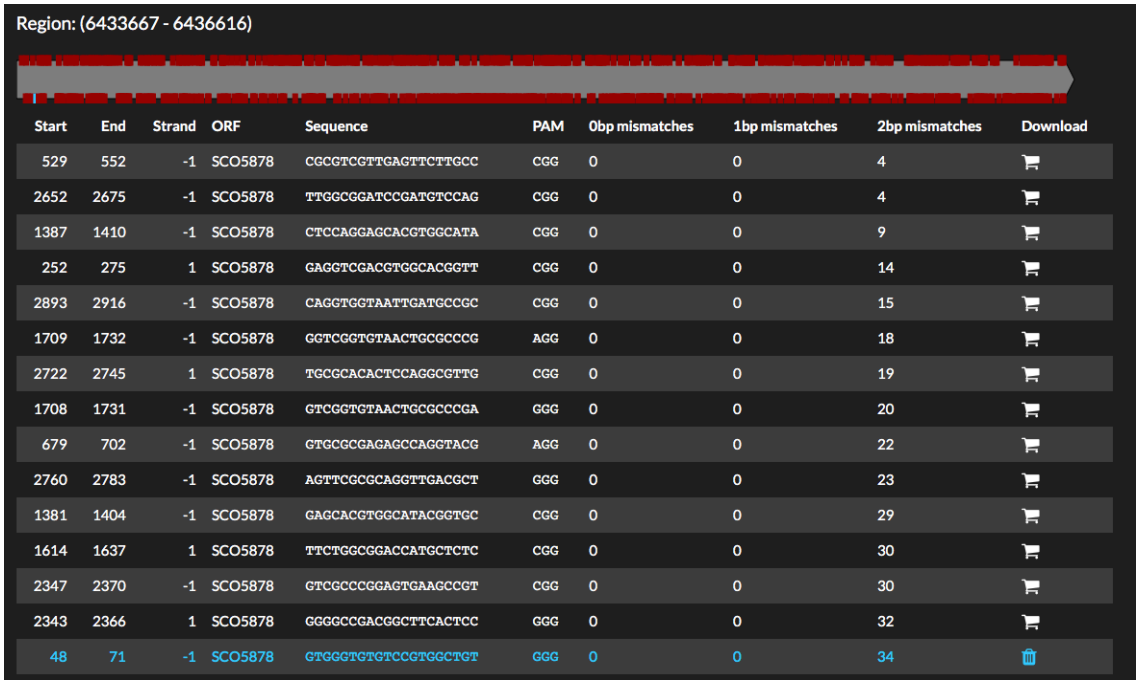
Screendumps of CRISPY-web[4] prediction of 20 bp protospacer sequences specific for **(I)** SCO5087, a core gene involved in actinorhodin (ACT) biosynthesis, and **(II)** SCO5878, a core gene of undecylprodigiosin (RED) biosynthesis, in *Streptomyces coelicolor* A3(2). The chosen 20 bp spacers are marked with blue on the schematic representation of the gene and in the list. For **(I)**, only the NT1 sequence appears in the list in the screenshot. The NT2 appears later in the list, but is still marked on the gene. NT1 binds at position 52 – 75 and NT2 at position 385 – 408 on the non-template strand of SCO5087. For **(II)**, one NT was chosen for SCO5878 as marked by the blue narrow square on the gene and the blue text in the list. This NT binds at position 48 – 71 on the non-template strand of SCO5878.

(I)

Region: (5529800 - 5531204)

Start	End	Strand	ORF	Sequence	PAM	0bp mismatches	1bp mismatches	2bp mismatches	Download
52	75	-1	SCO5087	GTGGCTCGAAGGAGGCTCGA	AGG	0	0	11	
29	52	1	SCO5087	GGCATCGAGGGGTCCCGTAT	CGG	0	0	15	
131	154	1	SCO5087	GAAGCGCAGAGTCGTCATCA	CGG	0	0	15	
188	211	-1	SCO5087	TGAGCAGTTCCCAGAACTGC	CGG	0	0	15	
251	274	1	SCO5087	CCCCTCGCCCTACCGTTCAC	AGG	0	0	17	
56	79	1	SCO5087	CGAGCCTCCTTCGAGCCACG	GGG	0	0	19	
473	496	1	SCO5087	GCGCGAGTATCTGCTGCTGT	CGG	0	0	20	
259	282	-1	SCO5087	CGCCGCGACCTGTGAACGGT	AGG	0	0	21	
55	78	1	SCO5087	TCGAGCCTCCTTCGAGCCAC	GGG	0	0	21	
935	958	1	SCO5087	CTGCAACGCGTACCACATGA	CGG	0	0	21	
910	933	-1	SCO5087	CGTCGCGTACCCCGAGATCT	CGG	0	0	23	
60	83	-1	SCO5087	TCGGCCCCGTGGCTCGAAGG	AGG	0	0	28	
467	490	-1	SCO5087	GCAGCAGATACTCGCGCTCC	AGG	0	0	35	
187	210	-1	SCO5087	GAGCAGTTCCCAGAACTGCC	GGG	0	0	43	
179	202	1	SCO5087	GAACGGCACCCGGCAGTTCT	GGG	0	0	50	
258	281	-1	SCO5087	GCCGCGACCTGTGAACGGTA	GGG	0	1	8	
54	77	1	SCO5087	TTCGAGCCTCCTTCGAGCCA	CGG	0	1	13	

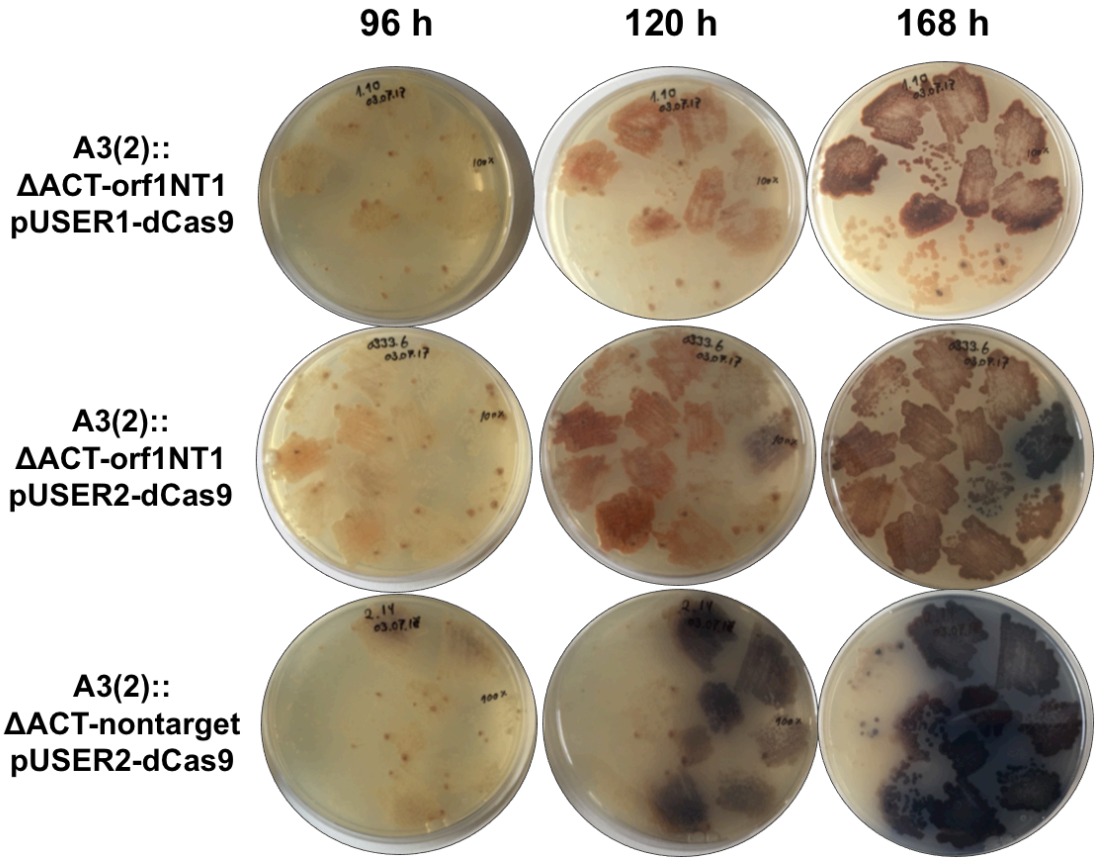
(II)


Region: (6433667 - 6436616)

Start	End	Strand	ORF	Sequence	PAM	0bp mismatches	1bp mismatches	2bp mismatches	Download
529	552	-1	SCO5878	CGCGTCGTTGAGTTCTTGCC	CGG	0	0	4	
2652	2675	-1	SCO5878	TTGGCGGATCCGATGTCCAG	CGG	0	0	4	
1387	1410	-1	SCO5878	CTCCAGGAGCACGTGGCATA	CGG	0	0	9	
252	275	1	SCO5878	GAGGTCGACGTGGCACGGTT	CGG	0	0	14	
2893	2916	-1	SCO5878	CAGGTGGTAATTGATGCCGC	CGG	0	0	15	
1709	1732	-1	SCO5878	GGTCGGTGTAACTGCGCCCG	AGG	0	0	18	
2722	2745	1	SCO5878	TGCGCACACTCCAGGCGTTG	CGG	0	0	19	
1708	1731	-1	SCO5878	GTCGGTGTAACTGCGCCCGA	GGG	0	0	20	
679	702	-1	SCO5878	GTGCGCGAGAGCCAGGTACG	AGG	0	0	22	
2760	2783	-1	SCO5878	AGTTCGCGCAGGTTGACGCT	GGG	0	0	23	
1381	1404	-1	SCO5878	GAGCACGTGGCATACGGTGC	CGG	0	0	29	
1614	1637	1	SCO5878	TTCTGGCGGACCATGCTCTC	CGG	0	0	30	
2347	2370	-1	SCO5878	GTCGCCCGGAGTGAAGCCGT	CGG	0	0	30	
2343	2366	1	SCO5878	GGGGCCGACGGCTTCACTCC	GGG	0	0	32	
48	71	-1	SCO5878	GTGGGTGTGTCCGTGGCTGT	GGG	0	0	34	

Figure S3: Actinorhodin (ACT) production over time for CRISPRi mutants

To verify successful CRISPRi silencing of core genes involved in actinorhodin (ACT) biosynthesis in *Streptomyces coelicolor* A3(2), plates of A3(2)::ΔACT-orf1NT1 (generated using either pUSER1-dCas9 or pUSER2-dCas9) and A3(2)::ΔACT-nontarget were visually examined at different time points. Pictures were taken at day 4 (96 h), day 5 (120 h), and day 7 (168 h), revealing the slight onset of actinorhodin (ACT) production at day 5. On day 7, plates were evaluated and it was concluded that both pUSER1-dCas9 and pUSER2-dCas9 can be used for gene repression of ACT production in *S. coelicolor* A3(2).

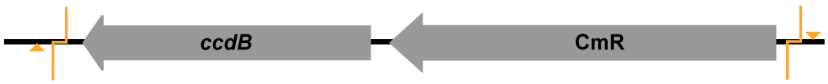


Note S1: USER cassettes (gBlock® gene fragments)

UC1 (gBlock® gene fragment)

Schematics

Original USER cassette (UC1) from Tong *et al.*[3]. This cassette contains the *ccdB* suicide gene with the chloramphenicol resistance marker (CmR) directly upstream. The cassette is removed from the pUSER1 constructs by a restriction and a nicking enzyme. The restriction (orange lines) and nicking (orange triangle) sites of the enzymes are marked in the fragment.



Sequence

GAATGCGTGCGATCGCAGCCTACTCGCTATTGTCCTCAATGCCGTATTAAATCATAAAAAGAAATAA
GAAAAAGAGGTGCGAGCCTCTTTTTTGTGTGACAAAATAAAAACATCTACCTATTCATATACGCTAG
TGTCATAGTCCTGAAAATCATCTGCATCAAGAACAATTTCACAACTCTTATACTTTTCTCTTACAAGT
CGTTCGGCTTCATCTGGATTTTCAGCCTCTATACTTACTAAACGTGATAAAGTTTCTGTAATTTCTACT
GTATCGACCTGCAGACTGGCTGTGTATAAGGGAGCCTGACATTTATATTCCCCAGAACATCAGGTTA
ATGGCGTTTTTGATGTCATTTTCGCGGTGGCTGAGATCAGCCACTTCTTCCCCGATAACGGAGACCGG
CACACTGGCCATATCGGTGGTCATCATGCGCCAGCTTTCATCCCCGATATGCACCACCGGGTAAAGT
TCACGGGAGACTTTATCTGACAGCAGACGTGCACTGGCCAGGGGGATCACCATCCGTCGCCCGGGCG
TGTCAATAATATCACTCTGTACATCCACAAACAGACGATAACGGCTCTCTCTTTTATAGGTGTAAACC
TTAAACTGCATTTCACCAGCCCCTGTTCTCGTCAGCAAAAGAGCCGTTCATTTCAATAAACCGGGCG
ACCTCAGCCATCCCTTCCTGATTTTCCGCTTTCCAGCGTTCGGCACGCAGACGACGGGCTTCATTCTG
CATGGTTGTGCTTACCAGACCGGAGATATTGACATCATATATGCCTTGAGCAACTGATAGCTGTCGCT
GTCAACTGTCACTGTAATACGCTGCTTCATAGCATACCTCTTTTTGACATACTTCGGGTATACATATC
AGTATATATTCTTATACCGCAAAAATCAGCGCGCAAATACGCATACTGTTATCTGGCTTTTAGTAAGC
CGGATCCACGCGGCGTTTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATT
CTGCCGACATGGAAGCCATCACAGACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGT
CGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTT
TAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCT
TTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCC
GGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTA
ACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGA
GCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGG
TCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAA
TGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTC
CATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTT
CATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAG
GGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTAT
TTATTCGGCGCAGCGATCGCGTGCATTC

UC2 (gBlock® gene fragment)

Schematics

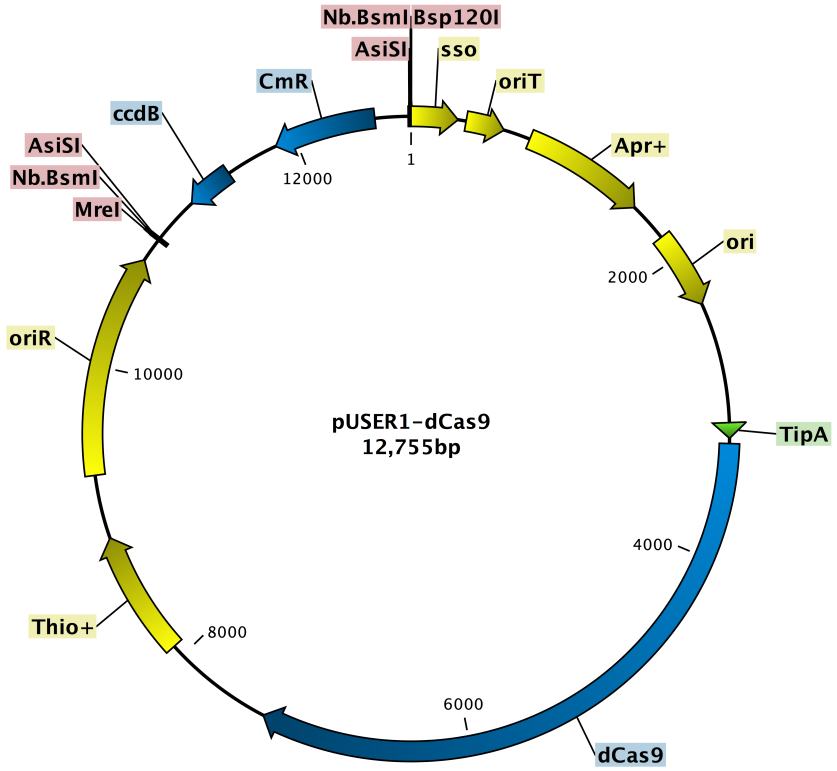
New USER cassette (UC2) ordered as a gBlock® gene fragment. UC2 contains the suicide gene *ccdB* with the chloramphenicol resistance marker (CmR) upstream. Additionally, the core sgRNA flanked by the *ermE** promoter and t_o terminator has been included in this fragment as well. The cassette is removed from the vector by a restriction and a nicking enzymes of which the restriction sites are marked in the gBlock® gene fragment (orange) below.

Sequence

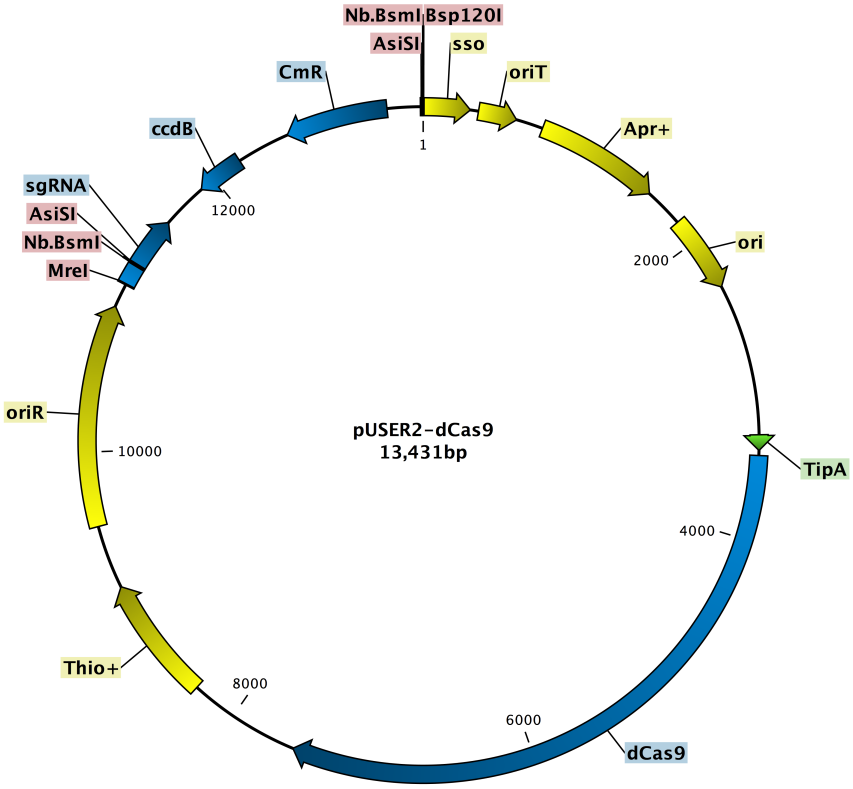
GCTAGCAAAGCGGTCGATCTTGACGGCTGGCGAGAGGTGCGGGGAGGATCTGACCGACGCGGTCCA
CACGTGGCACCGCGATGCTGTTGTGGGCACAATCGTGCCGGTTGGTAGGATCGACGGACTAGGAATG
CGTGCGATCGCGTGGCTCGAAGGAGGCTCGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT
AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTACCCTAGGAAAAGCTACGAT
GTTCCGGGGACTGCTGATCCGGTCAGCAGGTGGAAGAGGGACTGGATTCCAAAGTTCTCAATGCTGC
TTGCTGTTCTTGAATGGGGGGTCGTTGACGACGACATGGCTCGATTGGCGCGACAAGTTGCTGCGAT
TCTCACCAATAAAAAACGCCCGGCGGCAACGCAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGT
CACCTGAGGAGCCTACTCGCTATTGTCCTCAATGCCGTATTAAATCATAAAAAGAAATAAGAAAAAG
AGGTGCGAGCCTCTTTTTTGTGTGACAAAATAAAAACATCTACCTATTCATATACGCTAGTGTCATAG
TCCTGAAAATCATCTGCATCAAGAACAATTTCACAACTCTTATACTTTTCTCTTACAAGTCGTTCGGC
TTCATCTGGATTTTCAGCCTCTATACTTACTAAACGTGATAAAGTTTCTGTAATTTCTACTGTATCGAC
CTGCAGACTGGCTGTGTATAAGGGAGCCTGACATTTATATTCCCCAGAACATCAGGTTAATGGCGTTT
TTGATGTCATTTTCGCGGTGGCTGAGATCAGCCACTTCTTCCCCGATAACGGAGACCGGCACACTGG
CCATATCGGTGGTCATCATGCGCCAGCTTTCATCCCCGATATGCACCACCGGGTAAAGTTCACGGGA
GACTTTATCTGACAGCAGACGTGCACTGGCCAGGGGGATCACCATCCGTCGCCCGGGCGTGTCAATA
ATATCACTCTGTACATCCACAAACAGACGATAACGGCTCTCTCTTTTATAGGTGTAAACCTTAAACTG
CATTTCACCAGCCCCTGTTCTCGTCAGCAAAAGAGCCGTTCATTTCAATAAACCGGGCGACCTCAGC
CATCCCTTCCTGATTTTCCGCTTTCCAGCGTTCGGCACGCAGACGACGGGCTTCATTCTGCATGGTTG
TGCTTACCAGACCGGAGATATTGACATCATATATGCCTTGAGCAACTGATAGCTGTCGCTGTCAACTG
TCACTGTAATACGCTGCTTCATAGCATACCTCTTTTTGACATACTTCGGGTATACATATCAGTATATAT
TCTTATACCGCAAAAATCAGCGCGCAAATACGCATACTGTTATCTGGCTTTTAGTAAGCCGGATCCAC
GCGGCGTTTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACA
TGGAAGCCATCACAGACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCG
TATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAA
ACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAA
TAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGT
CGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTG
AACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCA
GGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAA
GGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAA
TGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCT
TCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTG
AAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCG
GTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGC
GCAGCGATCGCGTGCATTC

Note S2: Plasmid maps

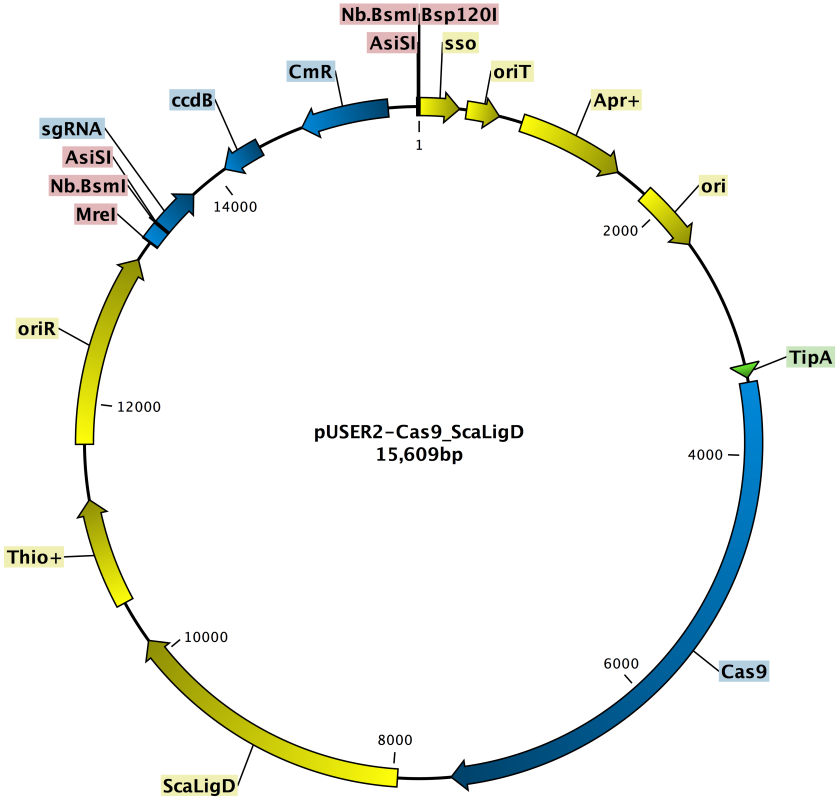
pUSER1-dCas9



pUSER2-dCas9



pUSER2-Cas9/ScaLigD



Supplementary References

1. MacNeil, D. J. *et al.* Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111,** 61–68 (1992).

2. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417,** 141–147 (2002).

3. Tong, Y., Robertsen, H. L., Blin, K., Weber, T. & Lee, S. Y. in *Methods in Molecular Biology* **1671,** 163–184 (2018).

4. Blin, K. CRISPy-web: An online resource to design sgRNAs for CRISPR applications. *Synth. Syst. Biotechnol.* **1,** 118–121 (2016).

Manuscript IV

Genetic perturbations caused by CRISPR-Cas9 genetic engineering of *Streptomyces collinus* Tü 365

Despite the advantages of implementing the CRISPR-Cas9 strategy for genome engineering in streptomycetes, the fact that the system is still in its infancy brings about certain limitations to its use. We experienced one such limitation of the technology when construction a *kirN*-deletion (Δ*kirN*) mutant of the kirromycin-producer strain *Streptomyces collinus* Tü 365. *kirN* encodes a crotonyl-CoA reductase/carboxylase (CCR), which is predicted to be responsible for provision of the ethylmalonyl-CoA extender units, incorporated at the C28 position in kirromycin (**Fig. 5**). CCRs are known to be involved in the assimilation of acetyl-CoA through an alternative glyoxylate pathway[82], and thus, we expected an isoenzyme of KirN from primary metabolism to partially complement its function. This should result in lowered kirromycin biosynthesis, but not complete loss of production in the Δ*kirN* mutant. However, to our surprise the CRISPR-Cas9-generated mutant of Δ*kirN* completely lost its ability to produce kirromycin (**Fig. 11**).

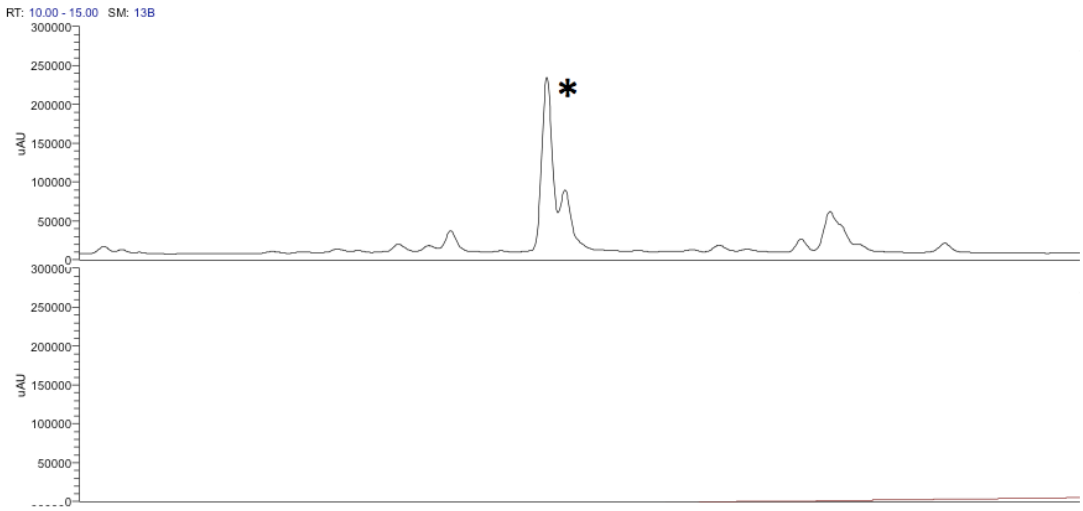


Figure 11 UV-Vis profiles of *Streptomyces collinus* Tü 365 wild type (top) and Δ*kirN* mutant (bottom). The asterisk marks kirromycin. No kirromycin-associated compounds were detected in the Δ*kirN* mutant.

Using Breseq[83] for whole genome sequence (WGS) analysis of the mutant, the results revealed an unexpected genotype, in which the CRISPR-Cas9 strategy had led to two large-scale deletions on both chromosomal arms of *S. collinus* Tü 365. The deletions of 0.79 Mb (left arm) and 0.63 Mb (right arm) explain why kirromycin production

was lost in the mutant, in that the BGC of kirromycin is present in two identical copies on the long terminal inverted repeats (TIRs) of the chromosomal arms[49,50] (see **Appendix 2 and 3**).

Several features characteristic for streptomycetes could provide hints to why this chromosome rearrangement was observed in the Δ*kirN* mutant. Chromosome replication in *Streptomyces* species is a bidirectional process starting from the centrally located *oriC*[84,85]. Core genes involved in primary metabolism are situated near the central region of the chromosome, whereas genes and gene clusters involved in secondary metabolism can be distributed across the chromosomes[86]. The natural habitat of streptomycetes can present a very competitive environment, in which the ability of the strain to quickly adapt to new growth conditions, including nutrient depletion or competing organism, is important for survival. Besides sporulation, the capability of streptomycetes to undergo recombination could provide a survival benefit in that exchange of genetic material, such as clusters encoding secondary metabolites, can result in mutants with novel phenotypes[87]. In addition, mutation rates have been reported to be inherently high (> 0.1 % mutants observed for colony-forming spores) in these organisms[88,89]. Chloramphenicol-sensitive (Cml^S) mutants of *Streptomyces lividans* 66 TK64 was found to arise at a 0.8 % average frequency in spores. This mutation in the Cml^S strains gave rise to increased instability, which resulted in 25 % arginine-auxothrophic mutants[90]. Spontaneous mutations have also been reported for *Streptomyces ambofaciens* DSM 40697, affecting its overall pigmentation with 17 % frequency. The pigment-deficient mutants displayed hypervariability, which could be further linked to large chromosomal deletions ranging from 250 – 2000 kb[88], all confined within the same genomic region.

The natural instability at the telomeres might provide hints to the observed genotype of the Δ*kirN* mutant. Several hypotheses exist trying to explain this instability, including the presence of transposon at the TIRs, which allow for intermolecular transposition between chromosomes and an exchange of the TIRs[91]. Alternatively, collapsed replication forks at single-strand DNA breaks are repaired by homologous recombination (HR). In the lack of suitable templates for HR, illegitimate recombination can result in deletions in or rearrangement of the chromosome[92,93].

Due to the unintended side effects of using the CRISPR-Cas9 technology for construction of genetic mutants in the duplicated kirromycin BGC of *S. collinus* Tü 365, an alternative strategy was undertaken to manipulate putative genes involved in kirromycin biosynthesis. Two of the mutants, Δ*kirOI* and Δ*kirOII*, included in our studies, was previously constructed by Kristina J. Laiple and David Worbs under the supervision of Tilmann Weber and Wolfgang Wohlleben at Tübingen University. The findings of our studies have resulted in **manuscript IV** presented in the following section.

Manuscript IV

Filling the Gaps in the Kirromycin Biosynthesis: Deciphering the Role of Genes Involved in Ethylmalonyl-CoA Supply and Tailoring Reactions

Helene L. Robertsen[a,1], Ewa M. Musiol-Kroll[a,2,3], Ling Ding[1], Kristina J. Laiple[2], Torben Hofeditz[4], Wolfgang Wohlleben[2,3], Sang Yup Lee[1,5], Stephanie Grond[4], and Tilmann Weber[1,*]

[a]Equal contribution

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, 2800 Kgs. Lyngby, Denmark

[2]Eberhard-Karls-Universität Tübingen, Interfakultäres Institut für Mikrobiologie und Infektionsmedizin, Microbiology/Biotechnology, Auf der Morgenstelle 28, 72076 Tübingen, Germany

[3]German Center for Infection Research (DZIF), Partner site Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany

[4]Eberhard-Karls-Universität Tübingen, Institute of Organic Chemistry, Auf der Morgenstelle 18, 72076 Tübingen, Germany

[5]Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

(Manuscript submitted to Scientific Reports on **November 30th 2017**)

Abstract

Kirromycin is the main product of the soil-dwelling *Streptomyces collinus* Tü 365. The elucidation of the biosynthetic pathway revealed that the antibiotic is synthesized via a unique combination of trans-/cis-AT type I polyketide synthases and non-ribosomal peptide synthetases (PKS I/NRPS). This was the first example of an assembly line integrating the three biosynthetic principles in one pathway. However, information about other enzymes involved in kirromycin biosynthesis remained scarce.

In this study, genes encoding tailoring enzymes KirM, KirHVI, KirOI, and KirOII, and the putative crotonyl-CoA reductase/carboxylase KirN were deleted, complemented, and the emerged products analysed. Derivatives were identified in mutants Δ*kirM*, Δ*kirHVI*, Δ*kirOI*, and Δ*kirOII*, of which the products of Δ*kirOII* and Δ*kirHVI* were subjected to NMR as well. Our efforts enabled functional assignment of the enzymes, demonstrating their involvement in kirromycin tailoring. In Δ*kirN*, the production of kirromycin was significantly decreased, which is advantageous for the production and purification of kirromycin analogues. The obtained data enabled us to clarify the putative roles of the studied enzymes, ultimately allowing us to fill many of the missing gaps in the biosynthesis of the complex antibiotic. Furthermore, this collection of mutants can serve as a toolbox for generation of new kirromycins.

Introduction

Streptomycetes are known for their profound ability to produce a diverse set of secondary metabolites with relevant pharmaceutical properties, including, but not limited to, antibiotics, anti-fungals, immunosuppressant, and antitumor agents. Polyketides comprise a noteworthy subset of these metabolites. They played and still play an important role in the clinic[1,2], hence highlighting the importance of their discovery.

Many complex polyketides are synthesized by megaenzyme complexes termed modular polyketide synthases (PKSs), which catalyse the controlled initiation and elongation of simple carboxylic acid monomers into polyketide chains. One example of a complex polyketide is the narrow-spectrum antibiotic kirromycin. The compound was first isolated from the soil-dwelling *Streptomyces collinus* Tü 365 in 1972[3]. The linear molecule is composed of three internal ring structures: a pyridone ring, a central tetrahydrofuran (THF) group, and a sugar-like moiety, also referred as goldinonic acid lactone[4] (**Fig. 1**). All three molecule structures are involved in the binding of the antibiotic kirromycin to the elongation factor (EF) Tu in prokaryotes[5,6]. This interaction prevents the conformational change between the GTP and GDP-bound form of the EF Tu, which results in stalling of protein biosynthesis. Although the structure and activity of kirromycin were known since the early 70s, most of the biosynthetic steps remained unexplored. First after the 82 kb kirromycin cluster (**Fig. 2**), encoding 28 genes, was identified and described[4], more details on the assembly of the compound were obtained. The hypothetical pathway of kirromycin indicated that the biosynthetic steps involve non-ribosomal peptide synthetases (NRPSs), acyltransferase (AT)-harbouring PKSs (*cis*-AT PKSs), AT-less PKSs (*trans*-AT PKSs), and several tailoring enzymes. This first description of a biosynthetic pathway involving three different types of multi-modular enzymes has made the assembly line a relevant model for studying complex *cis*-/*trans*-AT PKS/NRPS pathways.

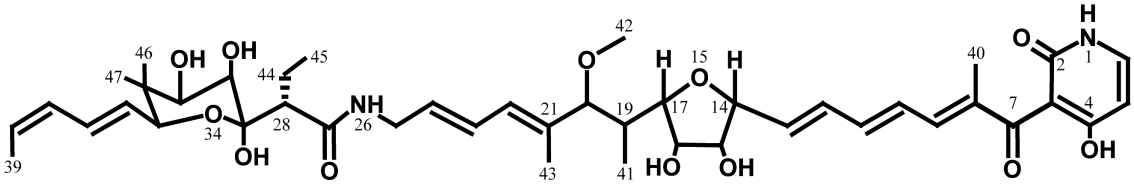


Figure 1 Structure of kirromycin. The molecule is composed of three intramolecular ring structures; a pyridone ring, a central tetrahydrofuran (THF) ring, and a sugar-like moiety termed goldinonic acid[6].

Directly upstream of the PKS/NRPS genes in the BGC of kirromycin, the Sfp-type phosphopantetheinyl transferase (PPTase), encoded by *kirP*, is found. The enzyme KirP catalyses the post-translational activation of the acyl carrier proteins (ACPs) and peptidyl carrier proteins (PCPs) and has a relaxed CoA substrate specificity, allowing it to accept both malonyl- and methylmalonyl-CoA[7]. The core PKS/NRPS enzymes in kirromycin are encoded by *kirAI – kirAVI* and *kirB*[4]. Closer examinations of the PKS-related genes and their products revealed KirAVI to be the only *cis*-AT PKS in the assembly line. The KirAVI PKS contains two AT domains (KirAVI-AT1 and KirAVI-AT2). ATs are responsible for the recognition, binding, and loading of extender units onto the acceptor proteins, the ACPs, in the PKS. According to the kirromycin structure and conserved amino acid (aa) sequence motives, the KirAVI-AT1 incorporates methylmalonate, while KirAVI-AT2 is responsible for introduction of malonate into the kirromycin polyketide backbone. The upstream PKSs, KirAI – KirAV, lack the integrated ATs. For these modules, extender units are provided by the discrete ATs[8]. The enzyme KirCI provides malonyl-CoA to several ACPs of the assembly line[9], whereas KirCII specifically recognizes only one ACP (KirAII-ACP5) and delivers the non-malonate precursor ethylmalonyl-CoA to module five, leading to the branched moiety at C28 of kirromycin[10]. Furthermore, it was demonstrated that KirCII loads both allyl- and propargyl-derived units onto its cognate ACP[11,12]. These features were exploited to generate kirromycin analogues with allyl and propargyl side chains *in vivo*[13]. The propargyl side chains can be utilized as educts for "click" chemistry reactions, which highlights the potential of the KirCII-ACP5 tool for the production of not only kirromycin derivatives, but also new polyketide products whenever the tool can be adapted to a heterologous assembly line.

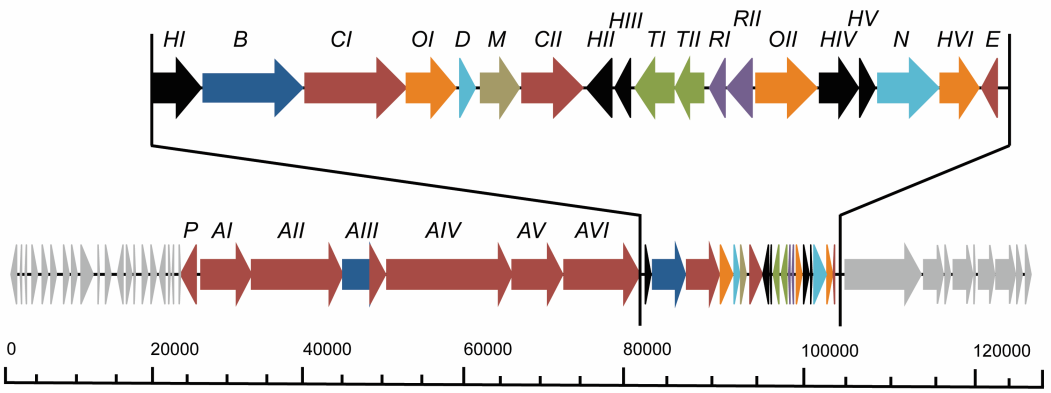


Figure 2 Schematic representation of the biosynthetic gene cluster encoding kirromycin. Black, hypothetical proteins; Blue, NRPS-related genes; Red, PKS-related genes; Orange, dehydrogenases and hydroxylases; Light blue, genes involved in precursor supply; Light brown, O-methyltransferase; Light green, transport-related genes; Purple, resistance-related genes; Light grey, genes putatively not involved in kirromycin biosynthesis (modified graphics from Weber *et al.*[4]).

kirB encodes the last core enzyme of the PKS I/NRPS assembly line of kirromycin, and has been found to be responsible for incorporation of the precursor β-alanine[4], which is biosynthesised by the aspartate-α-decarboxylase encoded by *kirD*[14]. Upon release from the assembly line this precursor is cyclized to form the pyridone ring[4]. Interestingly, no genes encoding a typical thioesterase (TE) were identified in the genome of the producer strain. Recently, Gui *et al.*[15] reported the discovery of a new group of Dieckmann cyclases involved in the biosynthesis of tetramic acid and pyridone scaffolds. A member of this new family of enzymes is also found in the kirromycin biosynthetic gene cluster (BGC). KirHI, previously assigned with a hypothetical function[4], was found to catalyse pyridone ring closure and the co-occurring release of the molecule chain from the PKS I/NRPS assembly line.

While the precursor loading of the PKS I/NRPS modules and the kirromycin polyketide chain elongation are well investigated (**Fig. 3A**), only limited knowledge about the provision of ethylmalonyl-CoA and the final tailoring reactions is available for this pathway. To date, the underlying reactions leading to closing of the THF ring, the introduction of hydroxyl groups at C16 and C30, methylation of the oxygen at O42, and the formation of the double bond between C5 and C6 in the pyridone ring have remained elusive. These enzymatic reactions cannot be explained from the enzymatic domains encoded within the PKS I/NRPS complex and thus, it was speculated that those product modifications are achieved through tailoring reactions.
In this study, the genes *kirM, kirHVI, kirOI, kirOII, kirHIV, kirHV* and *kirN*, which are encoded in the kirromycin BGC, but have no experimentally determined function, were examined by genetic inactivation and gene complementation. The compounds produced by the gene inactivation mutants were analysed and characterized by high performance liquid chromatography (HPLC)–high-resolution mass spectrometry (HRMS), and MS/MS. In addition, nuclear magnetic resonance (NMR) was employed for structural elucidation of the derivative produced by Δ*kirOII* and Δ*kirHVI.*

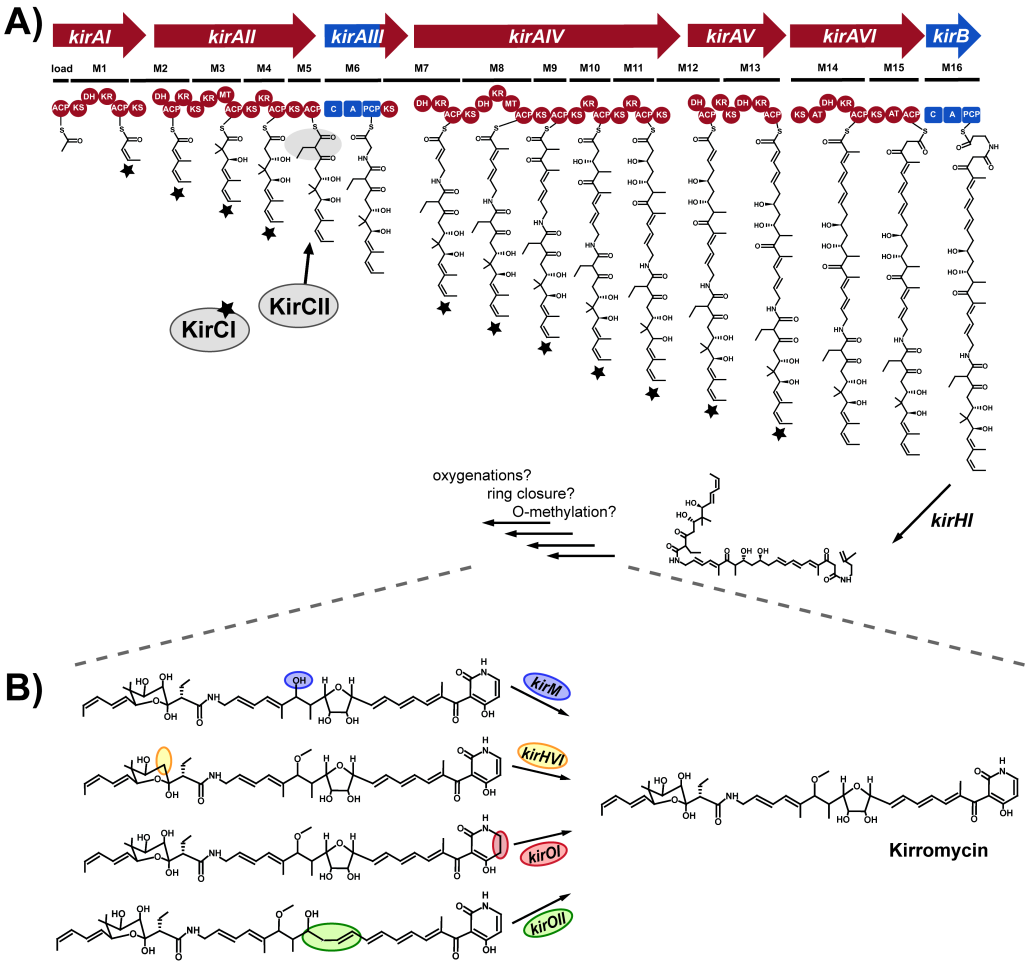


Figure 3 The hybrid PKS I/NRPS assembly line for kirromycin biosynthesis in *Streptomyces collinus* Tü 365. **A)** The mechanism behind the PKS I/NRPS chain elongation. Modified from Weber *et al.*[4]. B) Tailoring enzymes involved in the biosynthesis of kirromycin. ACP, acyl carrier protein; AT, acyltransferase; KS, ketosynthase; DH, dehydrogenase; KR, ketoreductase; ER, enoylreductase; MT, methyltransferase; C, condensation domain; A, adenylation domain; PCP, peptidyl carrier protein; KirCI/KirCII, trans-ATs; *kirHI*, Dieckmann cyclase; *kirM*, O-methyltransferase; *kirHVI*, phytanoyl-CoA dioxygenase; *kirOI/kirOII*, cytochrome P450-dependent hydroxylases.

Methods

Bacterial strains and general growth conditions

Plasmids and strains used in this study are listed in **Supplementary Table S1**.

Wild type strain *Streptomyces collinus* Tü 365 was obtained from the "Tübinger Stammsammlung". For routine cultivations of mutants, complemented mutants, and wild type strain, Tryptic Soy Broth (TSB) supplemented with nalidixic acid was used. Furthermore, mutants, into which the inactivated gene was introduced, were propagated in the presence of apramycin. Sporulation of wild type and mutant strains was performed on modified SFM agar (2 % mannitol, 2 % full fat soy flour, 20mM $MgCl_2$, 20mM $CaCl_2$, tap water). *Escherichia coli* DH5-α was used for standard cloning procedure of pGUSA21, pDrive, pA18mob, pJet1.2/blunt, pGM1190 and pRM4 constructs. Transformation was carried out according to the manufacturer's specifications and including a 1 h recovery step in Super Optimal broth with Catabolite repression (S.O.C.) medium. *E. coli* was grown at 37°C (200 rpm) in Luria Broth (LB) medium supplemented with appropriate antibiotics. In order to introduce the construct into streptomycetes, plasmid DNA was introduced into *E. coli* ET12567 (pUZ8002) by calcium chloride transformation. ET12567 strains were grown in LB supplemented with apramycin, chloramphenicol, and kanamycin for 16 hours, shaking at 200 rpm at 37°C.

Construction of *kirM, kirN, kirHIV* and *kirHV* gene inactivation plasmids

All primers used in this study are listed in **Supplementary Table S2**. A 2.1 kb cassette, containing pErmE* flanked by two 1 kb-fragments up- and downstream of the genes *kirM, kirHIV*, *kirHV,* and *kirN*, were amplified from pCRISPR-Cas9_USER- *kirM/kirHIV/kirHV/kirN*, which were constructed based on a previously described method[16] (see **Supplementary Table S1**). For the assembly, primers KP1 – KP8 and PCR programs listed in **Supplementary Table S2** were used. Each of the *kir* cassettes were assembled in pGUS A21 by Gibson cloning carried out at 50°C for 30 min. Correct clones of pHR1 – pHR4 were identified by control PCR and confirmed by Sanger sequencing with KP37/KP42.

Construction of *kirHVI* gene inactivation plasmid

To generate the *kirHVI* gene inactivation plasmid pHR2, 1 kb fragments, flanking the target gene, were amplified from the cosmid 2K05 DNA[4] using primers KP9/KP10 and KP13/KP14. pErmE* was amplified from plasmid pCRISPR-Cas9[17] with primers KP11/KP12. 1 M betaine was added to the PCRs to improve amplification of GC-rich stretches. Upon gel purification, the three fragments were assembled in pGUS A21[18] by Gibson cloning carried out at 50°C for 30 min, yielding pHR5. Correct clones of pHR5 were identified by control PCR and Sanger sequencing with KP37/KP38.

Construction of *kirOI* and *kirOII* gene inactivation plasmids

To construct pTL-*kirOI*, 1.2 kb of the flanking regions of the gene were amplified by PCR with primer pairs KP15/KP16 and KP17/KP18 using the PCR programs listed in **Supplementary Table S2**. For construction of pDW-*kirOII*, 2 kb of the flanking regions of *kirOII* were amplified by PCR with primers KP21/KP22 and KP23/KP24 using the PCR program listed in the **Supplementary Table S2**.

The left and right fragments, up- and downstream of *kirOI* and *kirOII,* were cloned into pA18mob resulting in the plasmids pTL-*kirOI* and pDW-*kirOII*, respectively. The thiostrepton resistance cassette (1.1 kb) was amplified from plasmid pSLE61[19] with primers KP19/KP20. The 1.1 kb fragment was first cloned into pDrive (Qiagen, Hilden, Germany) and then inserted via the *Xba*I sites into pTL-*kirOI* and pDW-*kirOII*, resulting in the plasmids pTL-*kirOI*-thio and pDW-*kirOII*-thio, respectively. The gene replacement mutants were confirmed by control PCRs targeting the apramycin and thiostrepton resistance cassettes with the primer pairs KP39/KP40 and KP19/KP20, respectively.

To distinguish between wild type and single/double crossover mutants, the primer pairs KP41/KP42 and KP43/KP44 were used for amplification of an internal region of (wild type) *kirOI/kirOII* or the thiostrepton resistance marker (single- or double crossover mutant).

In addition to PCR, Southern Blot analysis was used to confirm correct mutants of *kirOI* and *kirOII*. Here, the DIG-labelled probe for Southern Blot experiments was amplified from cosmid 1C24 DNA[4] with the primer pairs KP41/KP42 (*kirOI*) or KP43/KP44 (*kirOII*), using a 10X DIG DNA labelling mix (Roche, Mannheim, Germany). Genomic DNA preparation from the streptomycetes was performed with the NucleoSpin® Tissue Kit from Macherey-Nagel, Düren, Germany.

Intergeneric conjugation in *S. collinus* Tü 365

To introduce the gene inactivation plasmids into *S. collinus* Tü 365, a standard protocol for intergeneric conjugation was used[20]. Single crossover mutants were selected based on blue-white screening[18] in which spore dilution plates were overlaid with 20 mM X-Gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid). Control PCR was used to verify the single crossovers using primers KP72/KP73. For induction of the double crossover event, two days-old cultures of single crossover mutants, cultivated in TSB medium, were stressed for 24 h at 37°C, shaking at 160 rpm. Spore dilution plates were overlaid with 20 mM X-Gluc and white clones were picked for control PCRs with primers KP45 – KP58 (see **Supplementary Fig. S6**).

Construction of complementation plasmids for kirromycin *kir* mutants

kir genes were PCR amplified using the cosmid 2K05 (*kirM, kirHVI, kirOI, kirOII,* and *kirN*)[4] as template and the primers KP27 – KP36. 1M betaine was added to the PCRs to improve amplification of the GC-rich stretches. The DNA encoding the *kir* genes was purified from gel and cloned into blunt end pJet1.2 vector as instructed by the manufacturer. Correct clones were verified by control PCR and Sanger sequencing with primers KP59/KP60.

pJet1.2 clones of *kirM, kirHVI,* and *kirN* were subjected to restriction digestions with *NdeI* and *HindIII* together with DNA of integrative plasmid pRM4. The pJet1.2 clone of *kirOII* was subjected to restriction digestion with *NheI* and *HindIII* together with pRM4. Finally, pJet1.2-*kirOI* was subjected to restriction digestion with *HindIII* and *EcoRI* together with replicative plasmid pGM1190. The genes purified from gel and linearized vectors were ligated using T4 DNA Ligase and introduced into *E. coli* DH5-α competent cells according to manufacturer's instructions. The complementation plasmids of pGM1190 and pRM4 were verified by control PCR and Sanger sequencing with primers KP61/KP62 and KP63/KP64, respectively, and transferred to *S. collinus* mutant strains by standard intergeneric conjugation[20].

Kirromycin production assay, HPLC-HRMS, and MS/MS analyses

For the fermentations, 5 mL of two days-old precultures (grown in TSB), normalized according to weight, were used for inoculation of 95 mL kirromycin production medium composed of 1 % full-fat soy flour, 1 % D-mannitol, and 0.5 % $CaCO_3$ dissolved in tap water and pH adjusted to 7.4 prior to autoclaving. Fermentations were

carried out for six days at 30°C in a rotary shaker at 160 rpm. Cultures were extracted with 1:1 ethyl acetate for 90 minutes and evaporated in a Büchi® Rotavapor® RII evaporator equipped with jack and water bath. Dried extracts were redissolved in 500 µL methanol. HPLC–HRMS analysis was carried out with an Orbitrap Fusion connected to a Dionex Ultimate 3000 UHPLC pumping system (ThermoFisher Scientific, Waltham, MA, USA). UV-Vis detection was done using a DAD-3000 set to the range 200 – 600 nm. Samples were kept at 10.0°C in the autosampler during the analysis. 2 µL of each sample was injected into a C18 Acquity UPLC F5-3 HPLC column (2.1 × 100 mm, 1.8 µm,) at a flow rate of 0.4 *m*L/min, 30.0°C. Mobile phases A and B were 0.1 % formic acid in water and acetonitrile, respectively. Elution was done with a 30 min multistep system. After 5 % B for 1 min, a linear gradient started from 5 % B to 100 % B in 21 min, which was held for another 5 min and followed by re-equilibration to 5 % B until 30 min. Data was collected in both positive and negative ion modes with a scan range of (m/z) = 200 – 2000. MS/MS fragmentation was carried out using Ion Trap using collision-induced dissociation (CID) with the collision energy (30 %), RF Lens 60 %, AGC target 5.0e4, and scan range (m/z) 230 – 800. Data analyses were performed with the software Xcalibur 3.0.63 (Thermo Fisher Scientific Inc.).

NMR of kirromycin-HVI and kirromycin-OII derivatives

To confirm the structure of the kirromycin-OII derivative (**5**), a 2L-fermentation of the *ΔkirOII* mutant was carried out. Pure compound **5** (2.2 mg) was obtained by separation on a semi-preparative RP-18 column by HPLC. 1D and 2D NMR were acquired using standard pulse sequences on an 800 MHz Bruker Avance spectrometer with a 5 mm TCI cryo probe at NMR Center, Danish Technology University. CD_3OD was applied as solvent. Chemical shifts were usually expressed in parts per million (ppm, δ) relative to internal standard tetramethylsilane. In addition, kirromycin-HVI (**3**), kirromycin-OI (**4**), and **5** were analysed at the Institut für Organische Chemie in Tübingen University.

Results

Sequence analysis of genes and encoded products presumably involved in kirromycin biosynthesis in *S. collinus* Tü 365

The genes *kirM, kirHVI, kirOI, kirOII, kirHIV, kirHV* and *kirN* are located in the kirromycin gene cluster and thus a potential function of their gene products was postulated based on sequence analysis, using the sequence aligner DIAMOND against the MIBiG database.

For the tailoring enzymes, the closest homolog of KirM was the *O*-methyltransferase RapM (75 % aa identity and 87 % similarity), involved in rapamycin biosynthesis in *Streptomyces hygroscopicus* NRRL 5491[21]. With 31 % aa identity and 55 % similarity, the hydroxylase Fum3p involved in C5 hydroxylation in fumonisin biosynthesis in *Gibberella fujikuroi*[22] was the best hit for KirHVI. When the protein sequence of KirHVI was aligned against the Uniprot/Swissprot database, most of the hits were members of the phytanoyl-CoA dioxygenases enzyme family. KirOI and KirOII were putatively assigned to cytochrome P450 hydroxylases. KirOI displayed 49 % aa identity and 64 % similarity to the previously characterized cytochrome P450 hydroxylase TiaP2, responsible for hydroxylation of the C20 residue during Tiacumicin B biosynthesis[23]. For KirOII, the closest characterised homolog (40 % aa identity/59 % similarity) was found to be the 6-deoxyerythronolide_B hydroxylase encoded by *eryF*, which is involved in biosynthesis of erythromycin in *Saccharopolyspora erythraea*[24].

Additional proteins downstream of the PKS/NRPS biosynthetic core have remained assigned with hypothetical function, as it is the case for KirHIV and KirHV. The sequence analyses of KirHIV and KirHV, using DIAMOND alignment against the MIBiG database, did not yield any significant hits. Furthermore, based on the alignment against the UniProtKB/Swiss-Prot database, KirHIV and KirHV displayed similarity only to proteins with uncharacterized functions. Here, the best match for KirHIV (76 % aa identity and 88 % similarity) was an uncharacterized protein found in *Frankia* sp. BMG5.23. KirHV displayed the highest aa identity and similarity (71 and 80 %) to an uncharacterized protein from *Streptomyces* sp. CB00455. Hence, in respect of the sequence analyses, the functions of KirHIV and KirHV remained elusive.

The gene *kirN* is located downstream of the PKS I/NRPS-encoding region in the BGC (see **Supplementary Table S3**) and shows 76 % aa identity and 86 % similarity to the crotonyl-CoA reductase/carboxylase (CCR) SfaR from the sanglifehrin BGC[25]. In addition to this, the sequence analysis revealed that *kirN* displays between 70 to 76 % aa identity with several other CCRs found in other *Streptomyces* species.

Gene inactivations of *kirM, kirHVI, kirOI*, and *kirOII* result in production of kirromycin derivatives

To determine the role of the six genes presumed to be involved in tailoring reactions, two different gene replacement strategies were carried out. For the mutants in *kirM, kirHVI, kirHIV* and *kirHV*, the respective gene was replaced by the *ermE** promoter, whereas for *kirOI* and *kirOII,* a thiostrepton gene cassette together with the *ermE** promoter were used to replace the genes. Mutants were verified by control PCRs (see **Supplementary Fig. S6**), which resulted in either the expected bands or no product when the primers targeted internal regions of the deleted gene. The confirmed mutant clones were subjected to fermentation experiments carried out in parallel batches with wild type *S. collinus* Tü 365 as reference strain. The cultures were extracted with ethyl acetate, concentrated, and the extracts, including kirromycin and its derivatives, were dissolved in methanol. The samples were analysed by HRMS.

In the 30 min chromatographic method used for separation of kirromycin (**1**) and its derivatives, **1** appeared at 12.2 min with the main ion m/z = 795.4110 $[M-H]^-$ (**Fig. 4**). Furthermore, MS/MS fragmentation of the derivatives was carried out to gain insights into differences in the chemical formulas (see **Supplementary Fig. S2**). Kirromycin derivatives were detected in the mutants of Δ*kirM,* Δ*kirHVI,* Δ*kirOI,* and Δ*kirOII* (see **Fig. 4** and **Supplementary Table S3**). Gene inactivations of the hypothetical proteins encoded by *kirHIV* and *kirHV* did not affect the production of kirromycin and resulted in HPLC-HRMS data similar to the wild type (see **Supplementary Fig. S1-I** and **S1-II**). Thus, it was concluded that these proteins are not directly involved in kirromycin biosynthesis.

Gene inactivation of *kirM* resulted in the complete loss of production of **1**, and instead a new derivative, kirromycin-M (**2**), with the chemical formula $C_{42}H_{58}N_2O_{12}$ appeared in the chromatogram at 11.2 min with m/z = 781.3953 $[M-H]^-$ (**Fig. 4** and **Supplementary Table S3**). The MS/MS fragmentation of **2** gave rise to a fragment ion of m/z = 485.2292 $[M-H]^-$ (see **Supplementary Fig S2-II**). Compared to the

fragment ion of **1**, which was m/z = 499.2451 $[M-H]^-$, the difference of Δ14 corresponds to a difference of CH_2. This is in accordance with the predicted mass of a methyl group and gives strong evidence that KirM catalyses the *O*-methylation at the C20 position in the kirromycin precursor. The derivative produced by the Δ*kirM* mutant was thereby determined to be 20-O-demethyl-kirromycin, which for simplicity was referred to as **2**.

Similar to the Δ*kirM* mutant, the extracts of Δ*kirHVI,* Δ*kirOI*, and Δ*kirOII* were analysed using HPLC-HRMS and MS/MS. The chromatograms of these mutants revealed new prominent peaks, which were only detected with weak intensities in the extract of the wild type *S. collinus* Tü 365. Gene inactivation of *kirHVI* led to the production of kirromycin-HVI (**3**). Based on HRMS data (**Fig. 4** and **Supplementary Table S3)**, the chemical formula of **3** was calculated to be $C_{43}H_{60}N_2O_{11}$ and the corresponding ion m/z = 779.4177 $[M-H]^-$ appeared at 12.3 min. This derivative was already observed in the chromatogram of the WT. In this case, the presence of **3** in the wild type extract was concluded to be a shunt product from biosynthetic pathway and the result of incomplete biosynthesis of kirromycin. MS/MS fragmentation patterns for **1** and **3** yielded the same main fragment ion with m/z = 499.2457 $[M-H]^-$ (see **Supplementary Fig. S2-I** and **S2-III**). Because **1** is fragmented at the peptide bond of glycine at $C27^{13}$ and one of the detected peaks is that of the pyridine-containing fragment, this indicated that the sugar-containing fragment is missing a hydroxyl group on C30 of **3** in the Δ*kirHVI* mutant. Furthermore, ^1H- and ^{13}C-NMR corroborated the putative proposed structure of **3**, which could also be referred to as 30-deoxy-kirromycin (see **Supplementary Table S4** and **Fig. S4-IA/B**).

Gene inactivations of the putative P450-dependent hydroxylases encoded by *kirOI* and *kirOII* resulted in detection of the two derivatives kirromycin-OI (**4**) and kirromycin-OII (**5**) with m/z 797.4286 $[M-H]^-$ and 763.4213 $[M-H]^-$, which eluted at 13.6 min and 13.5 min, respectively (**Fig. 4** and **Supplementary Table S3**). Chemical formulas of **4** and **5** were calculated to be $C_{43}H_{62}N_2O_{12}$ and $C_{43}H_{60}N_2O_{10}$, respectively. This information together with the fragmentation data allowed for the assignment of KirOI to the double bond in the pyridone ring. Furthermore, the fragment ion of **4** of m/z = 501.26 $[M-H]^-$, which is the respective fragment of m/z = 499.2451 $[M-H]^-$ of **1**, showed a mass unit difference of Δ2 and could be assigned to the loss of two hydrogen atoms. The chemical name of **4** should therefore be 5,6-dihydro-kirromycin (see **Supplementary Fig. S2-I**, **S2-IV**, and **Supplementary Table S4**).

In the case of KirOII, the difference of two oxygen atoms underlined the hypothesis of the enzyme being involved in oxidation at C16 and presumably preparing the ring closure to the THF ring. MS/MS fragmentation of **5** yielded the fragment ion of m/z = 469.2548 $[M+H]^+$, which failed to give insight into the exact chemical structure of the derivative (see **Supplementary Fig. S2-V**). Therefore, NMR analyses, including ^{1}HNMR, ^{13}CNMR, COSY, HSQC, and HMBC were undertaken to gain more insight into the structure of this derivative (see **Supplementary Table S4, Supplementary Fig. S4-IIA/B**, and **Fig. S5**). To elucidate the structure of **5**, a 2 L-fermentation of the Δ*kirOII* mutant was carried out. Pure compound **5** (2.2 mg) was obtained by separation on a semi-preparative RP-18 column by HPLC. ^{13}CNMR spectrum showed the loss of signals for three oxygen-bearing carbons indicating opening of the furan ring and corroborated the chemical nomenclature 30-hydroxy-5,6-dehydro-1-N-demethyl-16-deoxy-kirromycin (see **Supplementary Table S4** and **Fig. S5-A**).

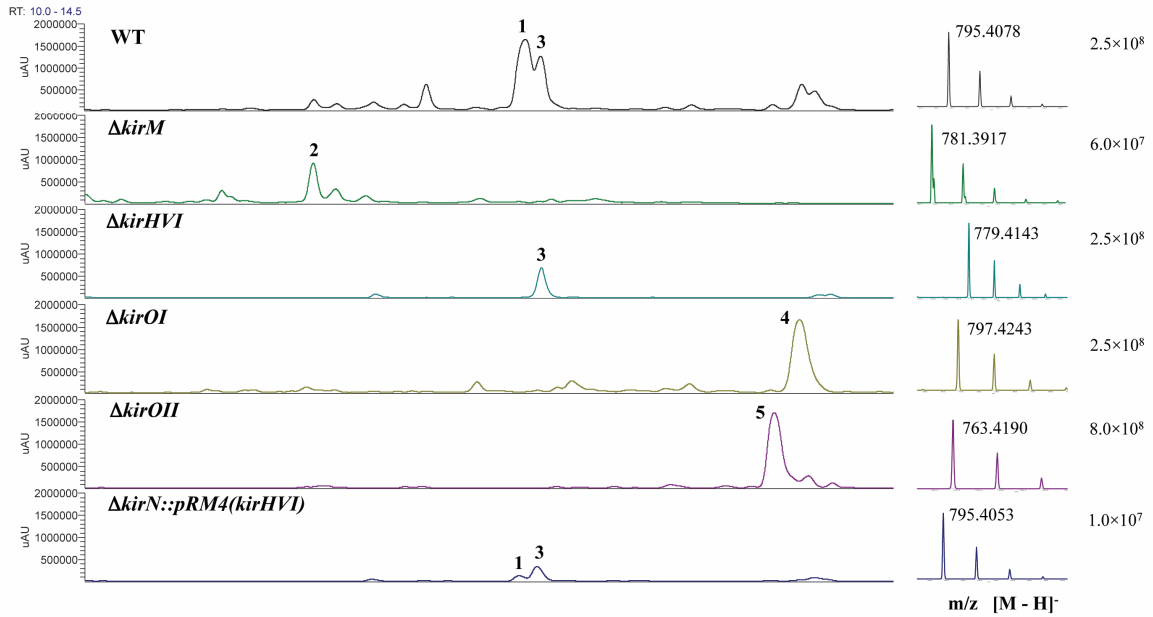


Figure 4 HPLC-ESI chromatograms and MS spectra for kirromycin, produced by the wild type strain, and derivatives produced by the mutants.

To rule out polar effects caused by disruption of the kirromycin BGC, gene complementations for each of the mutants, Δ*kirM,* Δ*kirHVI,* and Δ*kirOII*, were carried out using conjugation and the integrative plasmid pRM4, and for Δ*kirOI* using the replicative plasmid pGM1190 (see **Supplementary Fig. S3-I – S3-IV**). The complemented mutants were fermented in kirromycin production medium, extracted, and the extracts analysed by HPLC-HRMS. The complementation of the Δ*kirM* mutant resulted in similar kirromycin production levels compared to the wild type

(100 %). For the complementation of the Δ*kirHVI* mutant, approximately 30 % of the wild type levels of **1** could be restored, which was confirmed by both MS and UV-Vis quantification. The complementation of the Δ*kirOI* and Δ*kirOII* mutants resulted in only 10 and 20 % restored production of **1** compared to the wild type levels. Although there was no full complementation for these two mutants, the retention times and molecular masses, as determined by HRMS, clearly indicated that the observed effects were due to the inactivation of the individual genes and not due to polar effects on the biosynthetic assembly line.

Gene inactivation of *kirN* gives rise to lowered kirromycin production

The Δ*kirN* mutant and complemented mutant were constructed and tested using the same methods as described for the mutants above. The gene *kirN* encodes a putative CCR, which is postulated to provide the ethylmalonyl-CoA extender unit for the *trans*-AT KirCII, which then loads this substrate onto KirAII-ACP5. Unexpectedly, a deletion of *kirN* (Δ*kirN*) resulted in loss of kirromycin production. Instead the main derivative **3,** which was observed in the Δ*kirHVI* mutant*,* was detected. Genetic complementations with both pGM1190 and pRM4 of the Δ*kirN* mutant failed to restore production of **1**. However, upon gene complementation of the Δ*kirN* mutant with *kirHVI*, levels of **1** were partially restored to ~30 % of wild type levels. Since the replacement of *kirN* with the *ermE** promoter was expected to drive expression through the downstream gene *kirHVI*, we instead examined the genetic organisation. This revealed overlapping open reading frames (ORFs) of *kirN* and *kirHVI* and it was concluded that the genetic inactivation of *kirN* led to the disruption of downstream gene *kirHVI*. To test if production of **1** could be fully restored, pRM4 harbouring a copy of both *kirN* and *kirHVI* was constructed and used for complementation of Δ*kirN*. This resulted in a restoration of **1** in the Δ*kirN* mutant, reaching between 40 – 100 % of wild type levels (see **Supplementary Fig. S3-V** and **S3-VI**). Together, this data confirmed that the downstream gene *kirHVI* was affected by the gene inactivation of *kirN* and a full recovery of production of **1** was only possible when *kirN* and its neighbouring gene *kirHVI* were encoded on the complementation construct.

Discussion

In this study, the involvement of the seven genes, *kirM, kirHVI, kirOI, kirOII, kirHIV, kirHV,* and *kirN* in the biosynthesis of kirromycin in *S. collinus* Tü 365 was investigated experimentally based on gene inactivations, complementations, and product analyses by HPLC-HRMS, MS/MS, and NMR.

Our studies provide experimental evidence that KirM acts as an *O*-methyltransferase, giving rise to the methyl group on O42 in kirromycin. This finding is further strengthened by comparison of the role of the close homolog RapM in rapamycin biosynthesis *Streptomyces hygroscopicus* NRRL 5491. Gene complementation with *rapM* of the deletion mutant MG2-10 (Δ*rapKIJMNOQL*) resulted in the restored *O*-methylation at the C16 position in this molecule. Furthermore, it was found that altering the level of *O*-methylation in rapamycin resulted in the production of rapamycin analogues with altered properties, including increased solubility and improved Caco-2 permeability[21]. Although bioactivity assays have not yet been carried out for **2** in this study, it would be interesting to test if the missing methyl group results in a kirromycin derivative with altered pharmacokinetic properties. Such assays are planned for the future.

Based on the MS/MS (see **Supplementary Fig. S2-III**) and NMR analysis (see **Supplementary Table S4** and **S4-IA/B**), the main derivative **3** produced by Δ*kirHVI* lacks a hydroxyl group at position C30 in the kirromycin analogue. From the sequence-based alignment against the MIBiG database, KirHVI displays similarity to the hydroxylase Fum3p, which is responsible for attachment of a hydroxyl group at the C5 position in fumonisin B, produced by *Gibberella fujikuroi*[22]. Heterologous expression of Fum3p in *Saccharomyces cerevisiae* was previously conducted to clarify the enzymatic function of the protein and based on *in vitro* and *in vivo* findings it was possible to determine Fum3p to be a 2-ketoglutarate-dependent dioxygenase[26]. Since most protein hits from the UniProtKB/Swiss-Prot sequence alignment of KirHVI belonged to the family of phytanoyl Co-A dioxygenases it can be postulated that the hydroxylation catalysed by KirHVI in kirromycin biosynthesis occurs in a similar fashion (see **Supplementary Fig. S7-I**).

Furthermore, the function of KirOI and KirOII has been analysed in this study. The Δ*kirOI* mutant produces a kirromycin derivative, which is undoubtedly missing the double bond in the pyridone ring (see **Supplementary Table S4**). Here, the predicted

cytochrome P450 hydroxylase KirOI most likely forms a hydroxylated intermediate which then undergoes dehydrogenation giving rise the fully unsaturated pyridone ring (see **Supplementary Fig. S7-II**).

Our HRMS and MS/MS analyses failed to provide data suitable for clear structure prediction of the derivative **5** produced by the Δ*kirOII* mutant. Therefore, **5** was subjected to NMR studies, which enabled us to uncover a kirromycin analogue missing the THF ring structure and both hydroxyl groups at positions C15 and C16 (see **Supplementary Table S4, Fig. S4-IIA/B**, and **Fig. S5**). Based on our current hypothesis, the biosynthesis of kirromycin from **5** might be explained by a mechanism similar to that of the cytochrome P450 monooxygenase AurH, which is involved in conversion of deoxyaureothin to aureothin, through the formation of two C-O bonds, ultimately giving rise to a THF ring[27,28]. Similar to AurH, KirOII could introduce two hydroxyl groups at C14 and C16. Allylic substitution of the hydroxyl system could then lead to the formation of the THF ring system at the expense of H_2O. The alternative enzymatic pathway is found in **Supplementary Fig. S7-III**. In the future, enzymatic *in vitro* assays are necessary to fully establish the enzymatic reaction catalysed by KirOII.

Finally, we have studied the putative CCR encoded by *kirN*, which is expected to be involved in biosynthesis of the ethylmalonyl-CoA extender unit incorporated in kirromycin at the C28 position[4]. Studies on the involvement of CCRs in both primary and secondary metabolism have led to a revision of their function and today these enzymes are known for their involvement in acetyl-CoA assimilation through an alternative pathway distinct from the glyoxylate pathway[29,30]. Furthermore, sequence analysis of actinomycetes has revealed a tendency of co-clustering of CCRs with PKSs. This suggests an important role of the CCRs in polyketide derivatization in that they can facilitate the incorporation of polyketide extender units alternative to butyrate[30,31]. The putative role of KirN is furthered strengthened by the ^{13}CNMR studies on the N-methylated kirromycin analogue aurodox, in which Liu *et al.*[32,33] confirmed the incorporation of an intact butyrate unit at the C28 position in the molecule. The presence of an additional copy of the CCR, mainly involved in primary metabolism, in the genome of *S. collinus* Tü 365 gives rise to an inherent complementation of Δ*kirN* and, therefore, complete abolishment of kirromycin production was not expected. Surprisingly, the Δ*kirN* mutant generated in this study

gave rise to production of **3** instead of the expected lowered production of **1**. Examination of the genetic organization of the kirromycin BGC revealed the ORF of *kirN* to overlap with that of *kirHVI*, hence leading to disruption of the latter gene and production of **3**. Partial and full restoration of kirromycin production was achieved by genetic complementation of Δ*kirN* with *kirHVI* separately and with *kirN* and *kirHVI* together, respectively (see **Supplementary Fig. S3-V** and **S3-VI**). The observed phenomenon should be kept in mind for future BGC engineering efforts to account for polar effects in these often tightly regulated systems.

In our study, we demonstrated that the enzymes KirM, KirHVI, KirOI and KirOII play important roles in late stages of the biosynthesis of kirromycin. Derivatives of **1** were detected in all mutants, expect for Δ*kirN,* Δ*kirHIV,* and Δ*kirHV*. Although the genes *kirHIV* and *kirHV* are located in the kirromycin BGC, their inactivation had no effect on kirromycin biosynthesis and resulted in similar production profiles in both the wild type and mutant clones. Consequently, the data suggest that these hypothetical proteins are not directly involved in kirromycin biosynthesis. For the Δ*kirN* mutant, we observed a reduced production of **1**. While the Δ*kirM* mutant produces a derivative missing a methyl group at the O42 position, the main derivative produced by the Δ*kirHVI* mutant lacks the hydroxyl group at the C30 position in the goldinonic acid lactone. Finally, gene inactivations of the *kirOI* and *kirOII*, encoding the two cytochrome P450 hydroxylases, result in a missing double bond in the pyridine ring and no THF ring closure, respectively. These results and the new knowledge have allowed us to close some of the missing gaps in the biosynthetic pathway of kirromycin (**Fig. 3B**). Our studies does not allow for determining a defined sequence of the tailoring reactions. However, based on our findings, it could be argued that the enzymes act in an independent manner.

The better understanding of the complex kirromycin assembly line, which can be regarded as model for hybrid *cis*-/*trans*-AT PKS I/NRPSs, provides new incentive to the development of novel strategies for the production of polyketide derivatives, which might result in novel bioactive compounds. Our previous reports presented the use of the *trans*-AT KirCII, which accepts non-natural malonate-derived extender units, as a potential tool for polyketide engineering. The expression of a modified malonyl-CoA synthase T207G/M306I MatB and external feeding of the kirromycin producer strain with allyl- or propargylmalonic acid resulted in the production of

allyl- and propargyl-kirromycin[13]. However, the reported yields of propargyl-kirromycin were low compared to the production of the wild type kirromycin (22 % propargyl-kirromycin compared to 78 % kirromycin). In particular the Δ*kirN* mutant, generated and analysed in this study, will support the applicability of the MatB-KirCII-ACP5 system as the production of kirromycin is significantly reduced in this mutant. Hence, the use of the *ΔkirN* mutant as scaffold for production of kirromycin derivatives might result in improved yields and an easier accessibility of the desired compounds.

Acknowledgments

This work was supported by grants of the Novo Nordisk Foundation [to SYL, TW] and the German Center for Infection Biology (DZIF) [to WW]. We would like to thank Thomas Härtner and David Worbs for technical assistance and Kai Blin for the DIAMOND analysis. Also thanks to professor Charlotte Held Gotfredsen and Kasper Enemark-Rasmussen for their support in NMR spectra measurements for the kirromycin-OII derivative. We also thank Julian Saur for valuable discussion on biosynthesis.

Author Contributions

H.L.R. and E.M.M.-K. contributed equally to this work. E.M.M.-K. and T.W. planned the experiments, which were carried out by H.L.R. and K.L. (Δ*kirOI* mutant). L.D. conducted all HRMS analysis. L.D., S.G., and T.H. carried out and analysed the NMR for KirHVI and KirOII. H.L.R., E.M.M.-K., L.D., S.G., and T.W. wrote the paper, which was read, edited and approved by all authors.

Additional Information

The authors declare no conflict of interest.

References

1. Weissman, K. J. in *Methods in Enzymology* **459,** 3–16 (2009).

2. Butler, M. S. Natural products to drugs: natural product-derived compounds in clinical trials. *Nat. Prod. Rep.* **25,** 475–516 (2008).

3. Wolf, H. & Zähner, H. Stoffwechselprodukte von Mikroorganismen. *Arch. Mikrobiol.* **83,** 147–154 (1972).

4. Weber, T. *et al.* Molecular Analysis of the Kirromycin Biosynthetic Gene Cluster Revealed β-Alanine as Precursor of the Pyridone Moiety. *Chem. Biol.* **15,** 175–188 (2008).

5. Vogeley, L., Palm, G. J., Mesters, J. R. & Hilgenfeld, R. Conformational change of elongation factor Tu (EF-Tu) induced by antibiotic binding. Crystal structure of the complex between EF-Tu. GDP and aurodox. *J. Biol. Chem.* **276,** 17149–17155 (2001).

6. Olsthoorn-Tieleman, L. N., Palstra, R.-J. T. S., van Wezel, G. P., Bibb, M. J. & Pleij, C. W. A. Elongation factor Tu3 (EF-Tu3) from the kirromycin producer *Streptomyces ramocissimus* is resistant to three classes of EF-Tu-specific inhibitors. *J. Bacteriol.* **189,** 3581–3590 (2007).

7. Pavlidou, M. *et al.* The phosphopantetheinyl transferase KirP activates the ACP and PCP domains of the kirromycin NRPS/PKS of *Streptomyces collinus* Tü 365. *FEMS Microbiol. Lett.* **319,** 26–33 (2011).

8. Musiol, E. M. & Weber, T. Discrete acyltransferases involved in polyketide biosynthesis. *Medchemcomm* **3,** 871–886 (2012).

9. Musiol, E. M. *et al.* The AT(2) domain of KirCI loads malonyl extender units to the ACPs of the kirromycin PKS. *Chembiochem* **14,** 1343–1352 (2013).

10. Musiol, E. M. *et al.* Supramolecular Templating in Kirromycin Biosynthesis: The Acyltransferase KirCII Loads Ethylmalonyl-CoA Extender onto a Specific ACP of the *trans*-AT PKS. *Chem. Biol.* **18,** 438–444 (2011).

11. Koryakina, I. *et al.* Poly Specific *trans*-Acyltransferase Machinery Revealed via Engineered Acyl-CoA Synthetases. *ACS Chem. Biol.* **8,** 200–208 (2013).

12. Ye, Z., Musiol, E. M., Weber, T. & Williams, G. J. Reprogramming Acyl Carrier Protein Interactions of an Acyl-CoA Promiscuous trans-Acyltransferase. *Chem. Biol.* **21,** 636–646 (2014).

13. Musiol-Kroll, E. M. *et al.* Polyketide Bioderivatization Using the Promiscuous Acyltransferase KirCII. *ACS Synth. Biol.* **6,** 421–427 (2017).

14. Laiple, K. J., Härtner, T., Fiedler, H.-P., Wohlleben, W. & Weber, T. The kirromycin gene cluster of *Streptomyces collinus* Tü 365 codes for an aspartate-α-decarboxylase, KirD, which is involved in the biosynthesis of the precursor β-alanine. *J. Antibiot. (Tokyo).* **62,** 465–468 (2009).

15. Gui, C. *et al.* Discovery of a New Family of Dieckmann Cyclases Essential to Tetramic Acid and Pyridone-Based Natural Products Biosynthesis. *Org. Lett.* **17,** 628–631 (2015).

16. Tong, Y., Robertsen, H. L., Blin, K., Weber, T. & Lee, S. Y. in *Methods in Molecular Biology* **1671,** 163–184 (2018).

17. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* **4,** 1020–1029 (2015).

18. Myronovskyi, M., Welle, E., Fedorenko, V. & Luzhetskyy, A. Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. *Appl. Environ. Microbiol.* **77,** 5370–5383 (2011).

19. Muth, G., Wohlleben, W. & Puhler, A. The minimal replicon of the *Streptomyces ghanaensis* plasmid pSG5 identified by subcloning and Tn5 mutagenesis. *Mol. Gen. Genet.* **211,** 424–429 (1988).

20. Kieser, T. Practical *Streptomyces* genetics. (2000).

21. Gregory, M. A. *et al.* Rapamycin biosynthesis: elucidation of gene product function. *Org. Biomol. Chem.* **4,** 3565–3568 (2006).

22. Desjardins, A. E., Plattner, R. D. & Proctor, R. H. Linkage among genes responsible for fumonisin biosynthesis in *Gibberella fujikuroi* mating population A. *Appl. Environ. Microbiol.* **62,** 2571–2576 (1996).

23. Xiao, Y. *et al.* Characterization of Tiacumicin B Biosynthetic Gene Cluster Affording Diversified Tiacumicin Analogues and Revealing a Tailoring Dihalogenase. *J. Am. Chem. Soc.* **133,** 1092–1105 (2011).

24. Weber, J., Leung, J., Swanson, S., Idler, K. & McAlpine, J. An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. *Science (80-.).* **252,** 114–117 (1991).

25. Qu, X. *et al.* Cloning, sequencing and characterization of the biosynthetic gene cluster of sanglifehrin A, a potent cyclophilin inhibitor. *Mol. BioSyst.* **7,** 852–861 (2011).

26. Ding, Y., Bojja, R. S. & Du, L. Fum3p, a 2-Ketoglutarate-Dependent Dioxygenase Required for C-5 Hydroxylation of Fumonisins in *Fusarium verticillioides*. *Appl. Environ. Microbiol.* **70,** 1931–1934 (2004).

27. He, J. & Hertweck, C. Iteration as Programmed Event during Polyketide Assembly; Molecular Analysis of the Aureothin Biosynthesis Gene Cluster. *Chem. Biol.* **10,** 1225–1232 (2003).

28. Richter, M. E. A., Traitcheva, N., Knüpfer, U. & Hertweck, C. Sequential Asymmetric Polyketide Heterocyclization Catalyzed by a Single Cytochrome P450 Monooxygenase (AurH). *Angew. Chemie* **120,** 9004–9007 (2008).

29. Erb, T. J. *et al.* Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA

carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 10631–10636 (2007).

30. Wilson, M. C. & Moore, B. S. Beyond ethylmalonyl-CoA: The functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* **29,** 72–86 (2012).

31. Eustaquio, A. S. *et al.* Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. *Proc. Natl. Acad. Sci.* **106,** 12295–12300 (2009).

32. Liu, C. M., Maehr, H., Leach, M., Liu, M. & Miller, P. A. Biosynthesis of aurodox (antibiotic X-5108). incorporation of 14C-labelled precursors into aurodox. *J. Antibiot. (Tokyo).* **30,** 416–419 (1977).

33. Liu, C. M., Williams, T. H. & Pitcher, R. G. 13C-NMR studies on the biosynthesis of aurodox (antibiotic X-5108). *J. Antibiot. (Tokyo).* **32,** 414–417 (1979).

Tables

Table 1 Results of the sequence analysis of KirM, KirN, KirHVI, KirOI, KirOII, KirHIV, KirHV, and KirN

Gene	Inactivation plasmid	Complementation plasmid	nt/aa[1]	Highest similarity	MIBiG (1) / UniProt KB/Swiss-Prot (2)	Putative function	S/I (%)[2]	Closest homologue in
kirM	pHR1	pRM4-*kirM*	957/319	*rapM*	1	SAM-dependent *O*-methyltransferase	87/74	*Streptomyces hygroscopicus* NRRL 5491
kirHVI	pHR2	pRM4-*kirHVI*	846/282	*fum3p*	1	Hydroxylase	55/31	*Gibberella fujikuroi*
kirOI	pTL-*kirOI*-thio	pGM1190-*kirOI*	1200/400	*tiaP2*	1	Cytochrome P450 hydroxylase	64/49	*Dactylosporangium aurantiacum* subsp. *hamdenensis* NRRL 18085
kirOII	pDW-kirOII-thio	pRM4-*kirOII*	1218/406	*eryF*	1	Cytochrome P450 hydroxylase	59/40	*Saccharopolyspora erythraea* NRRL 2338
kirHIV	pHR3	-	783/261	A0A062 WMT4	2	Uncharacterised protein	-	*Frankia* sp. BMG5.23
kirHV	pHR4	-	396/132	A0A1Q5 MN63	2	Uncharacterised protein	-	*Streptomyces* sp. CB00455
kirN	pHR5	pRM4-*kirN+kirHVI*	1368/456	*sfaR*	1	Crotonyl-CoA reductase/carboxylase	86/76	*Streptomyces flaveolus* DSM 9954

[1]nt, nucleotides; aa, amino acid. [2]S, similarity; I, identity.

Supplementary Information

Filling the Gaps in the Kirromycin Biosynthesis: Deciphering the Role of Genes Involved in Ethylmalonyl-CoA Supply and Tailoring Reactions

Helene L. Robertsen[a,1], Ewa M. Musiol-Kroll[a,2,3], Ling Ding[1], Kristina J. Laiple[2], Torben Hofeditz[4], Wolfgang Wohlleben[2,3], Sang Yup Lee[1,5], Stephanie Grond[4], and Tilmann Weber[1,*]

[a]Equal contribution

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet 220, 2800 Kgs. Lyngby, Denmark

[2]Eberhard-Karls-Universität Tübingen, Interfakultäres Institut für Mikrobiologie und Infektionsmedizin, Microbiology/Biotechnology, Auf der Morgenstelle 28, 72076 Tübingen, Germany

[3]German Centre for Infection Research (DZIF), Partner site Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany

[4]Eberhard-Karls-Universität Tübingen, Institut für Organische Chemie, Auf der Morgenstelle 18, 72076 Tübingen, Germany

[5]Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

Table of Contents

Table S1: Strains and plasmids in this study

Source	Description/Genotype	Reference
	***E. coli* strains**	
E. coli DH5-α	*fhu*A2, Δ(*arg*F-*lac*Z)U169, *pho*A, *gln*V44, Φ80, Δ(*lac*Z)M15, *gyr*A96, *rec*A1, *rel*A1, *end*A1, *thi*-1, *hsd*R17	New England Biolabs
E. coli ET12567 (pUZ8002)	*dam*13::Tn9(ChlR), *dcm*-6, *hsd*M, *hsd*R, *rec*F143, *zjj*-201::Tn10, *gal*K2, *gal*T22, *ara*14, *lac*Y1, *xyl*-5, *leu*B6, *thi*-1, *ton*A31, *rps*L136, *his*G4, *tsx*-78, *mtl*I, *gln*V44. pUZ8002(KanR)	[1]
	***Streptomyces* strains**	
S. collinus Tü 365	Wild type *Streptomyces collinus* Tü 365	[2]
Δ*kirM*	*S. collinus* Tü 365 deletion mutant in *kirM*	This study
Δ*kirHVI*	*S. collinus* Tü 365 deletion mutant in *kirHVI*	This study
Δ*kirOI*	*S. collinus* gene replacement mutant in which *kirOI* is replaced by the thiostrepton resistance cassette with the *ermE** promoter downstream, ThioR	This study
Δ*kirOII*	*S. collinus* gene replacement mutant in which *kirOII* is replaced by the thiostrepton resistance cassette with the *ermE** promoter downstream, ThioR	This study
Δ*kirHIV*	*S. collinus* Tü 365 deletion mutant in *kirHIV*	This study
Δ*kirHV*	*S. collinus* Tü 365 deletion mutant in *kirHV*	This study
Δ*kirN*	*S. collinus* Tü 365 deletion mutant in *kirN*	This study
Δ*kirOI*-pGM1190	Δ*kirOI* complemented with replicative plasmid pGM1190-*kirOI*. ApraR, ThioR	This study
Δ*kirM*-pRM4	Δ*kirM* complemented with integrative plasmid pRM4-*kirM*. ApraR	This study
Δ*kirHVI*-pRM4	Δ*kirHVI* complemented with integrative plasmid pRM4-*kirHVI*. ApraR	This study
Δ*kirOII*-pRM4	Δ*kirOII* complemented with integrative plasmid pRM4-*kirOII*. ApraR, ThioR	This study
Δ*kirN*-pRM4_1	Δ*kirN* complemented with integrative plasmid pRM4-*kirHVI*. ApraR	This study
Δ*kirN*-pRM4_2	Δ*kirN* complemented with integrative plasmid pRM4-*kirN+HVI*. ApraR	This study
	Cosmids	
1C24	pOJ436 cosmid vector with a fragment of the genome of kirromycin producer *Streptomyces collinus* Tü 365	[3]
2K05	pOJ436 cosmid vector with a fragment of the genome of kirromycin producer *Streptomyces collinus* Tü 365	[3]
	Plasmids	
pJet1.2/blunt	*rep*(pMB1), *bla*(AmpR), *eco*47IR, PlacUV5, T7 RNA polymerase promoter. Part of the CloneJET™ PCR Cloning Kit used for high efficiency cloning of PCR products	Thermo Fisher Scientific Inc.
pDrive	T7 RNA polymerase promoter, LacZα, SP6 RNA polymerase promoter, AmpR, KanR, pUC origin, phage f1 origin of replication. For construction of Δ*kirOI* and Δ*kirOII* gene inactivation plasmids	Qiagen
pGUS A21	Promoter probe vector, pSETGUS with deleted KpnI fragment containing TipA promoter. *gusA*, Δ*int*, Δ*attB*, MCS from pUC21. Used for gene inactivation	[4]

pGM1190	pGM190 derivative in which the kanamycin resistance cassette is exchanged with the apramycin resistance gene *aac3*(3)IV. TipA promoter, ApraR, ThioR. Used for complementation of the *kir* mutant *kirOI*	G. Muth (personal communication)
pRM4	pSET152*ermEp* derivative with an artificial RBS, ΦC31 integration vector, pErmE*, ApraR. Used for complementation of the *kir* mutants	[5]
pA18mob	pK18mob derivative in which *aph*II is replaced with the apramycin resistance cassette. *ori*T, *lacZ*α, ApraR. Used for construction of gene replacement mutants of *kirOI* and *kirOII*.	[3]
pSLE61	pUC19 derivative. *lacZ*α, *ori*I, pSG5 replicon, AmpR, ThioR. For amplification of thiostrepton resistance cassette	[6]
pCRISPR-Cas9	Derivative of pGM1190 with Cas9 under the control of TipA promoter	[7]
pCRISPR-Cas9_USER-*ΔkirM*	USER-compatible destination vector, which is a derivative of pGM1190 with Cas9 under the control of TipA promoter. Cloned with pErmE* between sgRNA and 1 kb flanking regions right and left of *kirM* inserted in the USER cassette. ApraR	[7,8]
pCRISPR-Cas9_USER-*ΔkirHIV*	USER-compatible destination vector, which is a derivative of pGM1190 with Cas9 under the control of TipA promoter. Cloned with pErmE* between sgRNA and 1 kb flanking regions right and left of *kirHIV* inserted in the USER cassette. ApraR	[7,8]
pCRISPR-Cas9_USER-*ΔkirHV*	USER-compatible destination vector, which is a derivative of pGM1190 with Cas9 under the control of TipA promoter. Cloned with pErmE* between sgRNA and 1 kb flanking regions right and left of *kirHV* inserted in the USER cassette. ApraR	[7,8]
pCRISPR-Cas9_USER-*ΔkirN*	USER-compatible destination vector, which is a derivative of pGM1190 with Cas9 under the control of TipA promoter. Cloned with pErmE* between sgRNA and 1 kb flanking regions right and left of *kirN* inserted in the USER cassette. ApraR	[7,8]
pHR1	pGUS A21 deletion plasmid of *kirM* cloned with pErmE* between 1kb flanking regions of the gene. 2.1 kb construct cloned from *pCRISPR-Cas9*_USER-*kirM*. ApraR	This study
pHR2	pGUS A21 deletion plasmid of *kirHVI* cloned with pErmE* between 1kb flanking regions of the gene. ApraR	This study
pHR3	pGUS A21 deletion plasmid of *kirHIV* cloned with pErmE* between 1kb flanking regions of the gene. 2.1 kb construct cloned from pCRISPR-Cas9_USER-*kirHIV*. ApraR	This study
pHR4	pGUS A21 deletion plasmid of *kirHV* cloned with pErmE* between 1kb flanking regions of the gene. 2.1 kb construct cloned *from pCRISPR-Cas9* USER-*kirHV*. ApraR	This study
pHR5	pGUS A21 deletion plasmid of *kirN* cloned with pErmE* between 1kb flanking regions of the gene. 2.1 kb construct cloned from pCRISPR-Cas9_USER-*kirN*. ApraR	This study
pTL-*kirOI*	1.2 kb flanking regions left and right of *kirOI* cloned in pA18mob. ApraR	This study
pTL-*kirOI*-thio	pTL-*kirOI* cloned with thiostrepton resistance cassette inserted in *XbaI* site. ApraR, ThioR	This study
pDW-*kirOII*	2 kb flanking regions left and right of *kirOI* cloned in pA18mob. ApraR	This study
pDW-*kirOII*-thio	pDW-*kirOII* cloned with thiostrepton resistance cassette inserted in the *XbaI* site. ApraR, ThioR	This study
pJet1.2-*kirM*	pJet1.2/blunt with *kirM* cloned in the MCS. AmpR	This study

pJet1.2-*kirHVI*	pJet1.2/blunt with *kirHVI* cloned in the MCS. AmpR	This study
pJet1.2-*kirOI*	pJet1.2/blunt with *kirOI* cloned in the MCS. AmpR	This study
pJet1.2-*kirOII*	pJet1.2/blunt with *kirOII* cloned in the MCS. AmpR	This study
pJet1.2-*kirN*	pJet1.2/blunt with *kirN* cloned in the MCS. AmpR	This study
pJet1.2-*kirN+HVI*	pJet1.2/blunt with *kirN+HVI* cloned in the MCS. AmpR	This study
pGM1190-*kirOI*	Complementation plasmid of *kirOI* cloned in *HindIII* + *EcoRI* site	This study
pRM4-*kirM*	Complementation plasmid of *kirM* cloned in *NdeI* + *HindIII* site	This study
pRM4-*kirHVI*	Complementation plasmid of *kirHVI* cloned in *NdeI* + *HindIII* site	This study
pRM4-*kirOII*	Complementation plasmid of *kirOII* cloned in *NheI* + *HindIII* site	This study
pRM4-*kirN*	Complementation plasmid of *kirN* cloned in *NdeI* + HindIII site	This study
pRM-*kirN+HVI*	Complementation plasmid of *kirN+kirHVI* cloned in *NdeI* + *HindIII* site	This study

Table S2: Primers used in this study

ID	Primer name	Sequence (5' - 3')	PCR program
***kirM, kirN, kirHIV*, and *kirHV* inactivation mutants**			
KP1	kirM_pGUS_left_FW	CTGCAGACGCGTCGACGTCAAGC GACGAGGAGCTGGCG	1. 98°C, 0:30 min 2. 98°C, 0:10 min 3. 73°C, 0:30 min 4. 72°C, 1:00 min 5. 72°C, 5:00 min (Steps 2-4 × 34)
KP2	kirM_pGUS_left_RV	ATTACCCTGTTATCCCTAGTTAGT ACGCGGGCCCCAGG	
KP3	kirN_pGUS_left_FW	GGAGCTCGAATTCGAAGCTTCTG CAGACGCGTCGACGTCATATGCG ACCCCGACCAGGA	1. 98°C, 0:30 min 2. 98°C, 0:10 min 3. 75°C, 0:30 min 4. 72°C, 1:00 min 5. 72°C, 5:00 min (Steps 2-4 × 34)
KP4	kirN_pGUS_right_RV	GCGGGAAGCAGTGATAAGCATT ACCCTGTTATCCCTAGTTCTAGA ATCTCCGACGACGCGT	
KP5	kirHIV_pGUS_left_FW	GCAGACGCGTCGACGTCACTGCT CAAGTACTTCGACTCGAC	1. 98°C, 0:30 min 2. 98°C, 0:10 min 3. 70°C, 0:30 min 4. 72°C, 1:00 min 5. 72°C, 5:00 min (Steps 2-4 × 34)
KP6	kirHIV_pGUS_right_RV	ATTACCCTGTTATCCCTAGTTTGA CCAGCGCCAGTTGTG	
KP7	kirHV_pGUS_left_FW	CTGCAGACGCGTCGACGTCAGCT CAGCTTCGCCGC	1. 98°C, 0:30 min 2. 98°C, 0:10 min 3. 73°C, 0:30 min 4. 72°C, 1:00 min 5. 72°C, 5:00 min (Steps 2-4 × 34)
KP8	kirHV_pGUS_right_RV	ATTACCCTGTTATCCCTAGTTGA CGATGTCGGGATCCTCCC	
***kirHVI* inactivation mutant**			
KP9	kirHVI_pGUS_left_FW	GACGCGTCGACGTCACCGGAGTC GTGCTG	1. 98°C, 0:30 min 2. 98°C, 0:10 min 3. 72°C, 0:30 min 4. 72°C, 0:30 min 5. 72°C, 5:00 min (Steps 2-4 × 34)
KP10	kirHVI_pGUS_left_RV	GTCAAGATCGACCGCTTGTTTCC CTTCCCTG	
KP11	kirHVI_pGUS_ermE_FW	CTACAGGGAAGGGAAACAAGCG GTCGATCTTGACG	1. 98°C, 0:30 min 2. 98°C, 0:10 min 3. 74°C, 0:10 min 4. 72°C, 1:00 min 5. 72°C, 5:00 min (Steps 2-4 × 34)
KP12	kirHVI_pGUS_ermE_RV	CGACGGGCGGCTCGCCGTCGATC CTACCAA	
KP13	kirHVI_pGUS_right_FW	GGTAGGATCGACGGCGAGCCGC CCGTCG	1. 98°C, 0:30 min 2. 98°C, 0:10 min 3. 75°C, 0:30 min 4. 72°C, 0:30 min 5. 72°C, 5:00 min (Steps 2-4 × 34)
KP14	kirHVI_pGUS_right_RV	ATTACCCTGTTATCCCTAGTTGG ACGCGACCAGGCCGC	

***kirOI* and *kirOII* inactivation mutants**			
KP15	1ldOxy5	AACTGCAGTCAACATCACCTACGACC	1. 94°C, 2:00 min 2. 94°C, 1:15 min 3. 60°C, 1:30 min 4. 72°C, 1:30 min 5. 72°C, 5:00 min (Steps 2-4 × 30)
KP16	1ldOxy3	AATCTAGACTCGTCGTAGCCGGTG	
KP17	1rdOxy5	AATCTAGAGTCGGCCAGCAACTGG	
KP18	1rdOxy3	AAGAATTCGAACACGTCGATGTGCG	
KP19	thio-up	ACTCTAGATCACTGACGAATCGAGGTCGAGGA	
KP20	thio-low	AATCTAGAGGCGAATACTTCATATGCGGGGAT	
KP21	leftOIIeco5	GAATTCGTGCCAGGAGGTGATCC	1. 94°C, 2:00 min 2. 94°C, 1:15 min 3. 60°C, 1:30 min 4. 72°C, 2:00 min 5. 72°C, 5:00 min (Steps 2-4 × 30)
KP22	leftOIIxba3	TCTAGACTCGACGTCCGTCCAATC	
KP23	rightOIIxba5	TCTAGAGGCCCGCGCTTCCTGAAAG	
KP24	rightOIIhind3	AAGCTTCAGCGCGAACTGCGTCG	
Complementation plasmids			
KP27	kirN_NdeI_FW	AAACATATGCAAGACATCATCGACGCCG	1. 98°C, 0:30 min 2. 98°C, 0:10 min 3. 60°C, 0:30 min 4. 72°C, * 5. 72°C, 5:00 min (Steps 2-4 × 34)
KP28	kirN_HindIII_RV	AAAAAGCTTTCATTTGTTTCCCTTCCCTGTAGG	
KP29	kirM_NdeI_FW	AAACATATGAGCCAACCCGATGTGATGACC	
KP30	kirM_HindIII_RV	AAAAAGCTTTCAGACTCGGACGGCGAC	
KP31	kirHVI_NdeI_FW	AAACATATGACCGACGAAGACCTCGTCACG	
KP32	kirHVI_HindIII_RV	AAAAAGCTTTCAACCGGCGCGCTCGG	
KP33	kirOI_ HindIII_pGM1190_FW	AAAAAGCTTGTGTCCGAGACCGTTCGTCCCG	
KP34	kirOI_ EcoRI_pGM1190_RV	AAAGAATTCTCACCATGTCACCGGCAGCTCG	
KP35	kirOII_NheI_FW	AAAGCTAGCGTGACCGGAACATTCGATTGGACG	
KP36	kirOII_HindIII_RV	AAAAAGCTTTCACACGAGCTGGACGGGCAG	
Control PCRs and sequencing			
KP37	kirLEFT_ pGUS_seq_FW	GCTCGAATTCGAAGCTTCTGCAG	1. 95°C, 3:00 min 2. 95°C, 0:30 min 3. 61°C, 0:30 min 4. 72°C, 2:15 min 5. 72°C, 5:00 min (Steps 2-4 × 40)
KP38	kirRIGHT_ pGUS_seq_RV	GATAAGCATTACCCTGTTATCCCTAG	
KP39	apra-up	AGCTTCTCAACCTTGG	1. 94°C, 2:00 min 2. 94°C, 1:00 min 3. 50°C, 1:00 min 4. 72°C, 1:00 min 5. 72°C, 5:00 min (Steps 2-4 × 30)
KP40	apra-low	TCCGCCAAGGCAAAGC	
KP41	oxy1	GCAAAGGAGAGCGTTGTGTC	

KP42	oxy2	TGGTCTCCGTTCACCATGTC	1. 94°C, 2:00 min 2. 94°C, 1:00 min 3. 60°C, 1:00 min 4. 72°C, 1:10 min 5. 72°C, 5:00 min (Steps 2-4 × 30)
KP43	kirOII_check_FW	GCGATTGGGGATCTTGGTGA	
KP44	kirOII_check_RV	ATGGAAACGGTGGTCACACG	
KP45	kirM_int-500bp_FW	CCACCCTGTCCGGTCGCGG	1. 95°C, 3:00 min 2. 95°C, 0:30 min 3. 64°C, 0:30 min 4. 72°C, 2:30 min 5. 72°C, 5:00 min (Steps 2-4 × 40)
KP46	kirM_int-500bp_RV	GCATCGTCCGCAGCCGCTG	
KP47	kirN_int-500bp_FW	GTGTTCGCGTGGAACCCCGG	
KP48	kirN_int-500bp_RV	GGGATCCTCCCCACCCGTG	
KP49	kirHIV_int-500bp_FW	CCTGGTGGAGGCCGACAACG	
KP50	kirHIV_int-500bp_RV	CGCCTCGGCGAGTTCGGCG	
KP51	kirHV_int-500bp_FW	GCATGAACTCCGTGCCCCAGG	
KP52	kirHV_int-500bp_RV	GCCGCCCGGTCAGTGCACG	
KP53	kirHVI_int-500bp_FW	TGGAGGAGTTCCTGGACCCCG	
KP54	kirHVI_int-500bp_RV	CGCCGTGATACAGACCGCCG	
KP55	kirOI_int-500bp_FW	CGAGGAAGGCGGACCCACCG	
KP56	kirOI_int-500bp_RV	GCGTCTCCCTCGGTGAAGTGGTG	
KP57	kirOII_int-500bp_FW	GACGGACCTCGTCGCCGGG	
KP58	kirOII_int-500bp_RV	GAGCTGGATCTGGCCGTCCTTG	
KP59	pJet_check_FW	CGACTCACTATAGGGAGAGCGGC	1. 95°C, 3:00 min 2. 95°C, 0:30 min 3. 60°C, 0:30 min 4. 72°C, ** 5. 72°C, 5:00 min (Steps 2-4 × 40)
KP60	pJet_check_RV	AAGAACATCGATTTTCCATGGCAG	
KP61	pGM1190_seq_FW	CACGCGGAACGTCCGGG	
KP62	pGM1190_seq_RV	CCGCTGAAACTGTTGAAAGTTGTTTAGC	
KP63	pRM4_seq_FW	CAGTCACGACGTTGTAAAACGACGG	
KP64	pRM4_seq_RV	GGCACCGCGATGCTGTTGTG	

* 0:15 min for *kirHV*. 0:30 min for *kirM/kirHIV/kirHVI*. 0:45 min for *kirN/kirOI/kirOII*.

** 1:00 min for *kirM/kirHVI*. 1:30 min for *kirN/kirOI/kirOII*.

Table S3: HPLC-ESI-HRMS data for kirromycin and derivatives

Compound	RT (min)	m/z for [M-H$^-$]	Chemical Formula	MS/MS fragment (m/z [M-H$^-$])	MS/MS fragment (m/z [M+H$^+$])
Kirromycin (**1**)*	12.2	795.4110	$C_{43}H_{60}N_2O_{12}$	499.2451	501.2578
Kirromycin-M (**2**)	11.2	781.3953	$C_{42}H_{58}N_2O_{12}$	485.2292	487.2292
Kirromycin-HVI (**3**)	12.3	779.4177	$C_{43}H_{60}N_2O_{11}$	499.2457	501.2586
Kirromycin-OI (**4**)	13.6	797.4286	$C_{43}H_{62}N_2O_{12}$	501.26	503.2768
Kirromycin-OII (**5**)	13.5	763.4213	$C_{43}H_{60}N_2O_{10}$	467.2548	469.2698

*Δ*kirN*(*kirHVI*-complemented), Δ*kirHIV,* and Δ*kirHV* have similar chromatograms and MS profiles to those of the wild type kirromycin (**1**), albeit production is lowered in Δ*kirN*(*kirHVI*-complemented).

Table S4: ^{1}HNMR spectra and assignments

^{1}H NMR data (**A**) and ^{13}C NMR data (**B**) for 5,6-dihydro-kirromycin (**4** from Δ*kirOI*), 30-hydroxy-5,6-dehydro-1-*N*-desmethyl-16-deoxy-kirrothricin (**5**, from Δ*kirOII*), and 30-deoxy-kirromycin (**3** from Δ*kirHVI*) in CD_3OD at resonance frequency [1)]600 resp. 150.8 MHz or [2)]700 resp. 176.1 MHz. The chemical shifts, obtained from nuclear magnetic resonance (NMR) spectroscopy analysis, are represented by Δ ppm.

	5,6-dihydro-kirromycin (**4,** from Δ*kirOI*)		**30-hydroxy-5,6-dehydro-1-N-demethyl-16-deoxy-kirrothricin** (**5**, from Δ*kirOII*)		**30-deoxy-kirromycin** (**3,** from Δ*kirHVI*)	
Position	**δ_H [ppm]** (Integral, Type, J in Hz)[1)]	**δ_C [ppm]**	**δ_H [ppm]** (Integral, Type, J in Hz)[2)]	**δ_C [ppm]**	**δ_H [ppm]** (Integral, Type, J in Hz)[2)]	**δ_C [ppm]**
1	-	-	-	-		-
2	-	n.d.	-	164.5		164.5
3	-	79.5[00]	-	112.72 112.64	-	112.5
4	-	169.8	-	169.18, 169.01	-	169.4
5	2.54, 1H, dd, *J* = 7.0 (br)	37.3 (br)[#]	6.12, 1H, d, *J* = 7.4	102.0	6.11, 1H, s, br	102.1
6	3.42, 1H, m, *J* = 7.0 (br)	37.9	7.40, 1H, d, *J* = 7.4	137.8	7.39, 1H, br	137.8
7	-	n.d.	-	198.55, 198.63	-	198.9
8	-	138.8[00]	-	142.9	-	141.4
9	6.20, m	134.3[00]	6.97, 1H, ddd, *J* = 11.4, 5.3, 1.2	143.3	6.94, 1H, br	142.5
10	6.56, m[00]	129.3	6.73, 1H, dd, *J* = 11.4, 14.7	129.5	6.69, 1H, b, dd, *J* = 12.5, 12.5	129.4
11	6.56, m[00]	137.2[00]	6.67, 1H, d, *J* = 14.7, 10.7	142.3	6.07, 1H, dd, *J* = 15.1, 7.2	137.0
12	6.40, 1H, *J* = 11.5, 15.0	133.9 (br)	6.61, 1H, dd, *J* = 14.7, 10.7	142.5	6.45, 1H, dd, *J* = 15.0, 11.0	133.6
13	5.99, m	135.1 (br)	6.46, 1H, dd, *J* = 10.5, 5.9 6.44, 1H, dd, *J* = 10.6, 5.7	138.7 134.0	6.52, 1H, m (with 24)	136.1
14	4.25, 1H, dd, *J* = 12.5	81.8	6.37, 0.6 H, dd, *J* = 14.7, 10.5	132.1	4.31, 1H, m	81.6
15	4.28, 1H, dd, *J* = 12.5	75.0	5.90, 0.4 H, dt, *J* = 14.9, 7.4; 6.26, 0.6H dd, *J* = 14.8, 10.4 (also 6.21, dd, *J* = 11.3, 11.3, 5.67, m,; 6.78, m)	136.02 133.50 (131.1, 132.2, 133.5)	4.32, 1H, m	75.0
16	4.17, 1H, dd, *J* = 4.4	74.0	2.25, m	40.0,	4.19, 1H, dd, *J* = 4.4	74.0

	5,6-dihydro-kirromycin (**4,** from Δ*kirOI*)		**30-hydroxy-5,6-dehydro-1-N-demethyl-16-deoxy-kirrothricin** (**5**, from Δ*kirOII*)		**30-deoxy-kirromycin** (**3,** from Δ*kirHVI*)	
Position	**δ_H [ppm]** (Integral, Type, J in Hz)[1)]	**δ_C [ppm]**	**δ_H [ppm]** (Integral, Type, J in Hz)[2)]	**δ_C [ppm]**	**δ_H [ppm]** (Integral, Type, J in Hz)[2)]	**δ_C [ppm]**
			2.40, m 2.43, m	40.7, 34.8		
17	3.68, 1H, dd, *J* = 4.0, 7.5	84.9	4.06, 1H, m	70.9	3.71, 1H, dd, 7.3, 4.1	85.0
18	-	-	-	-	-	-
19	2.16, 1H, ddq, *J* = 2.0, 7.0	36.8	1.69, 1H, m	40.8	2.20, 1H, m	36.8
20	3.32, s	91.9	3.46, 1H, d, *J* = 10.1	90.0	3.34, 1H, d, 9.2	91.9
21	-	136.0	-	136.7	-	136.1
22	5.96, 1H, d (br), *J* = 10	130.9	5.99, 1H, d, *J* = 10.5	130.4	5.69, 1H, dd, *J* = 15.0, 6.0	130.8
23	6.49, 1H, m, *J* = 11.0, 16.0	128.3	6.51, 1H, dd, *J* = 15.2, 11.0	128.3	6.56, 1H, dd, *J* = 15.2, 11.0	128.0
24	5.67, 1H, m, *J* = 6.0, 16.0	131.0	5.66, 1H, m (with 35)	130.6	6.53, 1H, dd, *J* = 15.2, 11.0	128.2
25a	3.86, 1H, dd, *J* = 6.5, 15.0	42.0	3.89, 1H, dd, *J* = 15.8, 5.5	42.1	3.90, 1H, dd, *J* = 15.8, 5.8	42.0
25b	3.93, 1H, dd, *J* = 6.5, 15.0		3.95, 1H, dd, *J* = 15.8, 6.7		3.95, 1H, dd, *J* = 15.8, 5.9	
26-N	-	-	-	-	-	-
27	-	177.9	-	177.9	-	176.9
28	2.82, 1H, dd, *J* = 4.0, 11.0	52.5	2.84, 1H, dd, *J* = 11.0, 4.5	52.5	2.39, 1H, dd, *J* = 11.1, 4.2	58.3
29	-	101.0	-	101.0	-	99.5
30	3.65, 1H, d, *J* = 4.0	71.4	3.68, 1H, d, *J* = 3.6	71.4	1.95, 1H, dd, *J* = 12.6, 4.6 1.48, 1H, dd, *J* = 12.6, 12.5	38.3
31	3.57, 1H, d, *J* = 4.0	74.0	3.60, 1H, d, *J* = 3.6	74.0	3.74, 1H, d, *J* = 12.5, 4.6	72.6
32	-	40.0	-	39.9	-	40.3
33	4.21, 1H, d, *J* = 6.0	77.3	4.24, 1H, t, *J* = 5.8	77.3	4.17, 1H, d, *J* = 6.0	77.7
34-O	-	-			-	-

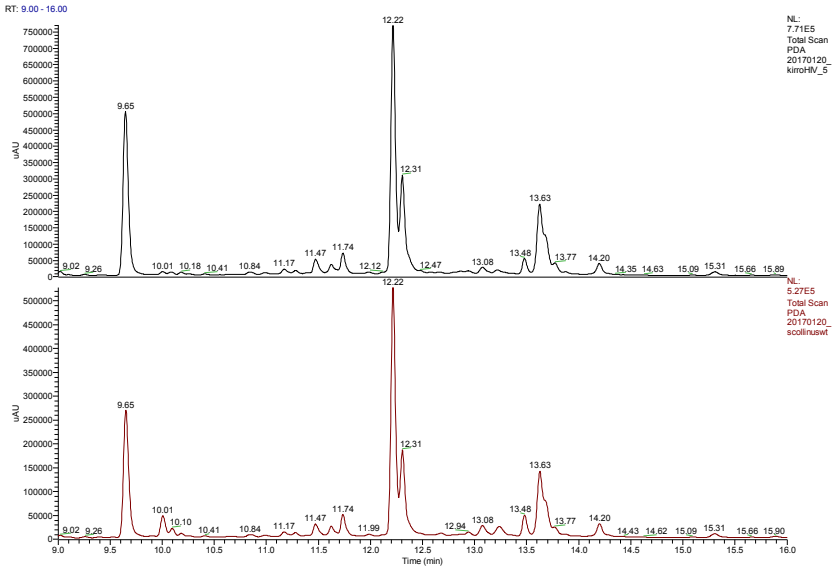
	5,6-dihydro-kirromycin (**4,** from Δ*kirOI*)		**30-hydroxy-5,6-dehydro-1-N-demethyl-16-deoxy-kirrothricin** (**5**, from Δ*kirOII*)		**30-deoxy-kirromycin** (**3,** from Δ*kirHVI*)	
Position	**δ_H [ppm]** (Integral, Type, J in Hz)[1)]	**δ_C [ppm]**	**δ_H [ppm]** (Integral, Type, J in Hz)[2)]	**δ_C [ppm]**	**δ_H [ppm]** (Integral, Type, J in Hz)[2)]	**δ_C [ppm]**
35	5.62, 1H, dd, *J* = 6.0, 15.0	130.5	5.66, 1H, m (with 24)	130.6	5.64, 1H, m (with 24)	130.3
36	6.54, m	127.7	6.58, 1H, dd, *J* = 14.8, 11.8	127.7	6.55, 1H, m (with 24)	127.9
37	5.99, m	130.6	6.01, 1H, ddd, *J* = 10.8, 10.8, 1.5	130.5	6.01, 1H, ddd, *J* = 10.8, 10.8, 1.0	130.4
38	5.44, 1H, dq, *J* = 7.0, 11.0	126.4	5.49, 1H, dq, *J* = 11.0, 7.0	126.4	5.49, 1H, dq, *J* = 11.0, 6.9	126.5
39	1.73, 3H, dd, *J* = 2.0, 7.0	13.7	1.77, 3H, dm, *J* = 7.0	13.7	1.77, 3H, dd, *J* = 7.1, 1.5	13.7
40	1.95, s (br)	14.4 (br)	2.00, 2.04, 3H, d, *J* = 0.9	11.61, 11.68	2.01, 3H,	11.7
41	0.81, 1H, d, *J* = 7.0	13.9	0.69, 3H, d, *J* = 7.0	9.49, 9.57	0.84, 3H, d, *J* = 6.8	13.9
42	3.15, 3H, s (sh)	56.2	3.17, 3H, s	56.32	3.18, 3H, s	56.2
43	1.66, 3H, s (sh)	11.2	1.63, 3H, s	11.0 (sh)	1.69, 3H, s	11.2
44	1.70, m	21.1	1.73, 2H, m	21.1	1.73, 2H, m	21.6
45	0.92, 3H, t, *J* = 7.5	12.3	0.95, 3H, t, *J* = 7.4	12.3	0.95, 3H, t, *J* = 7.3	12.4
46	0.89, 3H, s[00]	15.9	0.92, 3H, s	15.9	0.77, 3H, s	12.4
47	0.89, 3H, s[00]	24.6	0.92, 3H, s	24.6	0.94, 3H, s	23.0

[#]: assigned due to HSQC correlations. [00]: not unambiguously assigned due to tautomerization, missing 2D correlations, or overlaying signals. n.d.: not detected. n.a.: not assigned.

Figure S1-I – S1-II: UV-VIS profiles of Δ*kirHIV* and Δ*kirHV* mutants

The Δ*kirHIV* and Δ*kirHV* mutants were grown in parallel with *Streptomyces collinus* Tü 365 (wild type) in kirromycin production medium and extracted with ethyl acetate. The extracts were dried, redissolved and analysed using HPLC-ESI-HRMS. The recorded UV profiles of Δ*kirHIV* **(I)** and Δ*kirHV* **(II)** (black chromatograms) were similar to those of the wild type (red chromatograms). These results indicate that the two genes *kirHIV* and *kirHV* are not directly involved in biosynthesis of kirromycin. In the case of Δ*kirHV* **(II)**, we suspect that the lower intensity of the peak at 12.3 min is due to an almost complete assembly of kirromycin molecules and with that fewer shunt products.

(I)



(II)

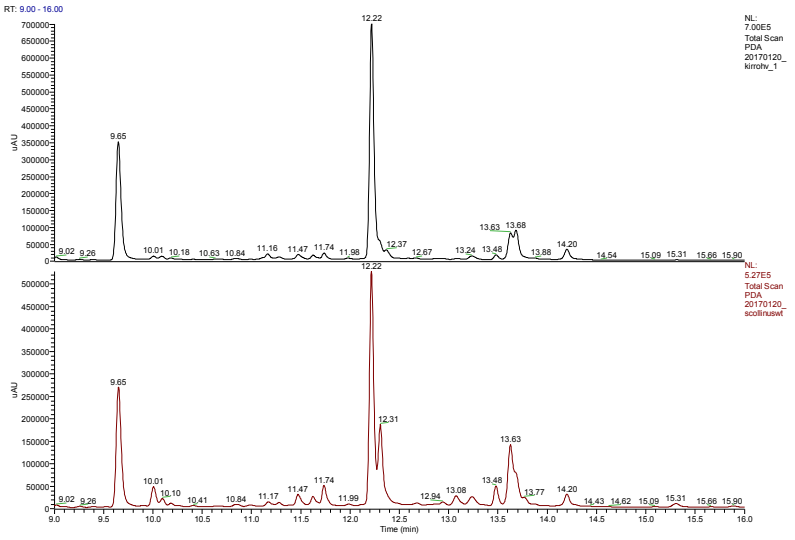
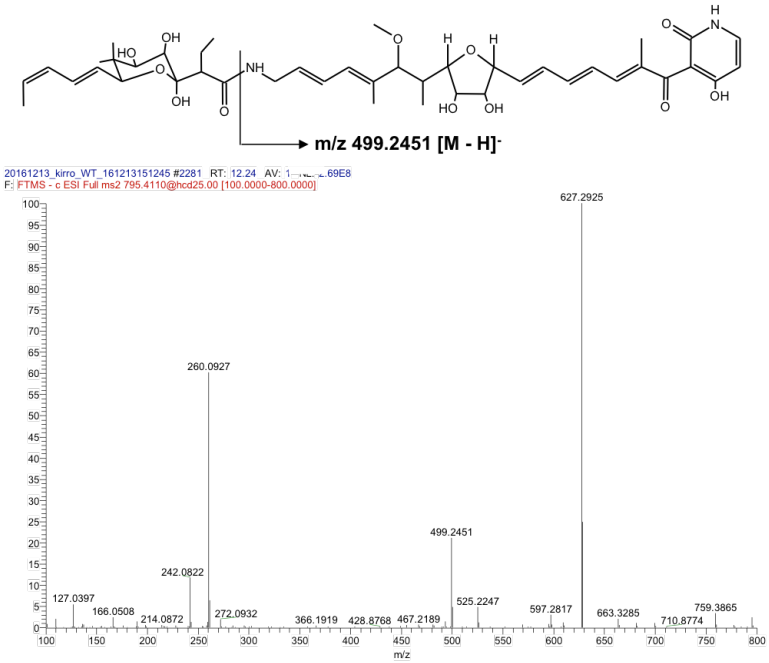


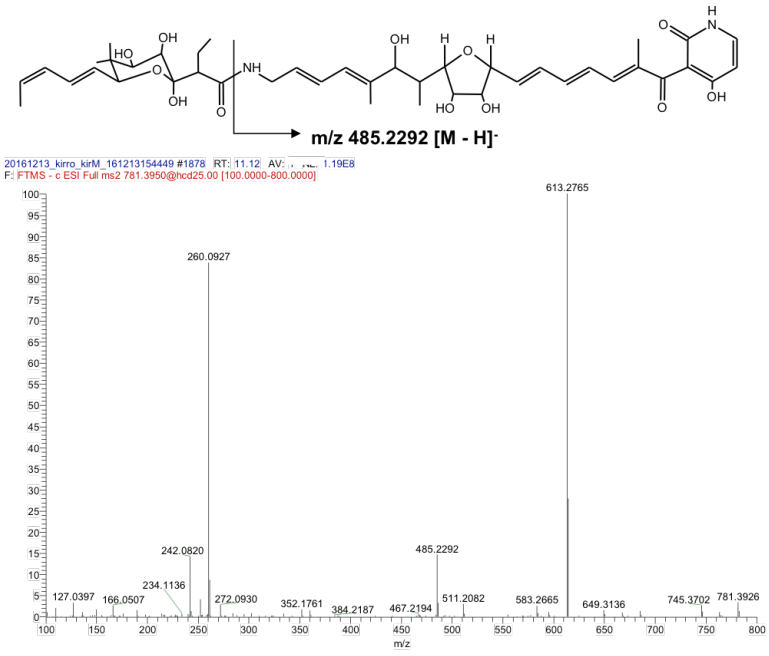
Figure S2-I – S2-V: MS/MS spectra of kirromycin and derivatives

MS/MS spectra for **(I)** kirromycin (**1**), **(II)** kirromycin-M (**2**), **(III)** kirromycin-HVI (**3**), **(IV)** kirromycin-OI (**4**), and **(V)** kirromycin-OII (**5**). Recorded in negative mode (ESI-) with higher-energy collisional dissociation (HCD) of 25 eV for **1**, **2**, **3**, and **5** and collision-induced dissociation (CID) of 30eV for **4**.

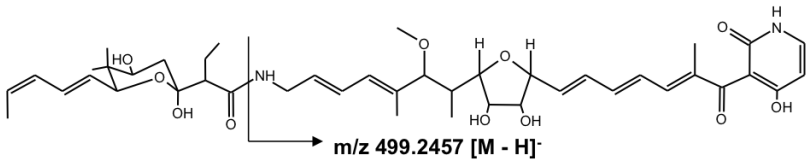
(I)

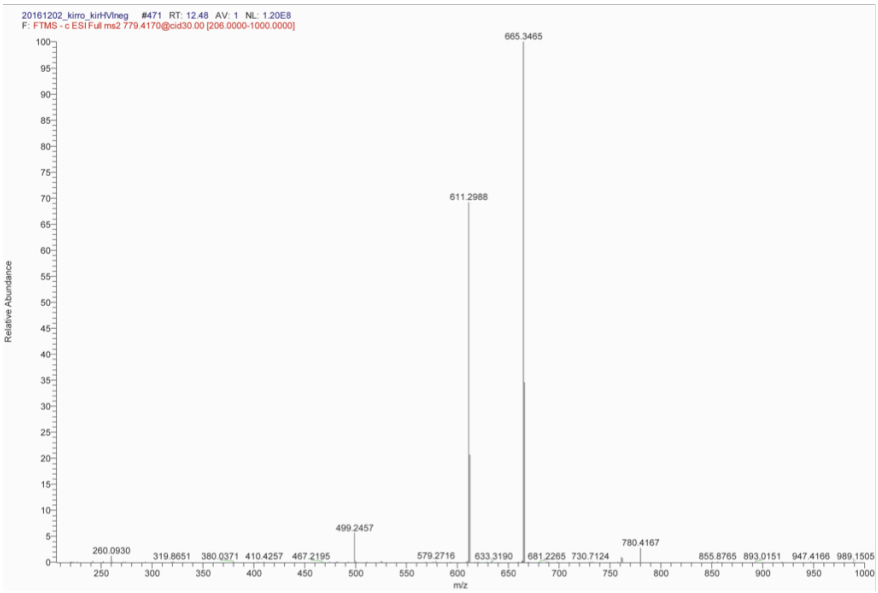


(II)

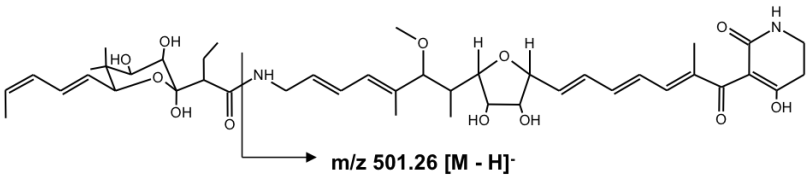


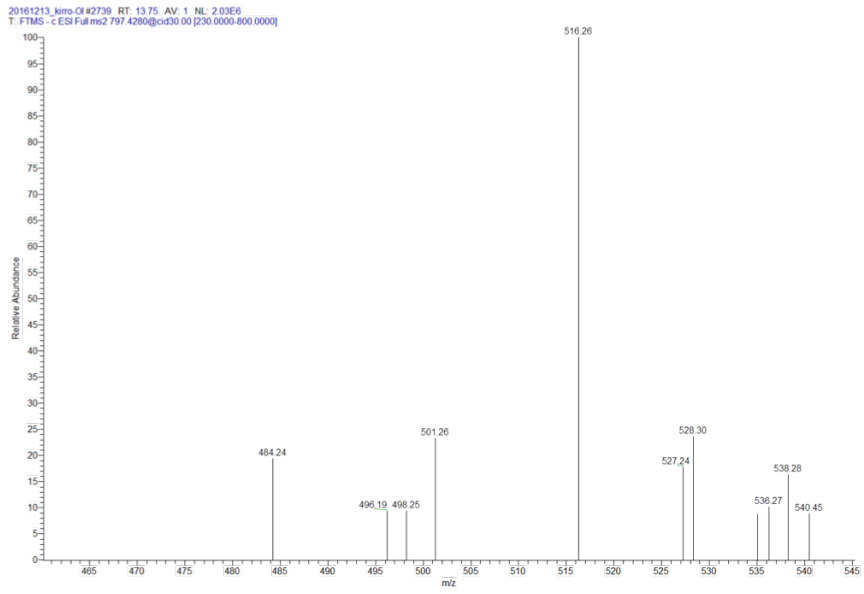
(III)


m/z 499.2457 [M - H]-

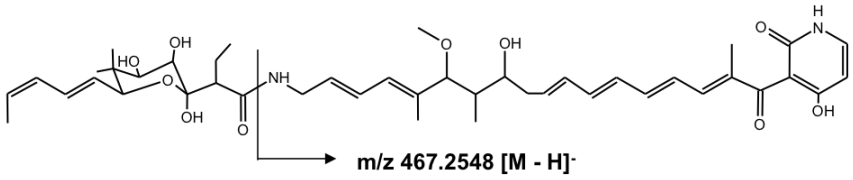

20161202_kirro_kirHVlneg #471 RT: 12.48 AV: 1 NL: 1.20E8
F: FTMS - c ESI Full ms2 779.4170@cid30.00 [206.0000-1000.0000]
Relative Abundance
m/z
665.3465
611.2988
499.2457
780.4167
260.0930
319.8651
380.0371
410.4257
467.2195
579.2716
633.3190
681.2265
730.7124
855.8765
893.0151
947.4166
989.1505

(IV)


m/z 501.26 [M - H]-


20161213_kirro-OI #2739 RT: 13.75 AV: 1 NL: 2.03E6
T: FTMS - c ESI Full ms2 797.4280@cid30.00 [230.0000-800.0000]
Relative Abundance
m/z
516.26
528.30
501.26
484.24
527.24
538.28
536.27
540.45
496.19
498.25

(V)


OH
HO
O
OH
O
NH
O
OH
O
H
N
O
OH
O
m/z 467.2548 [M - H]-

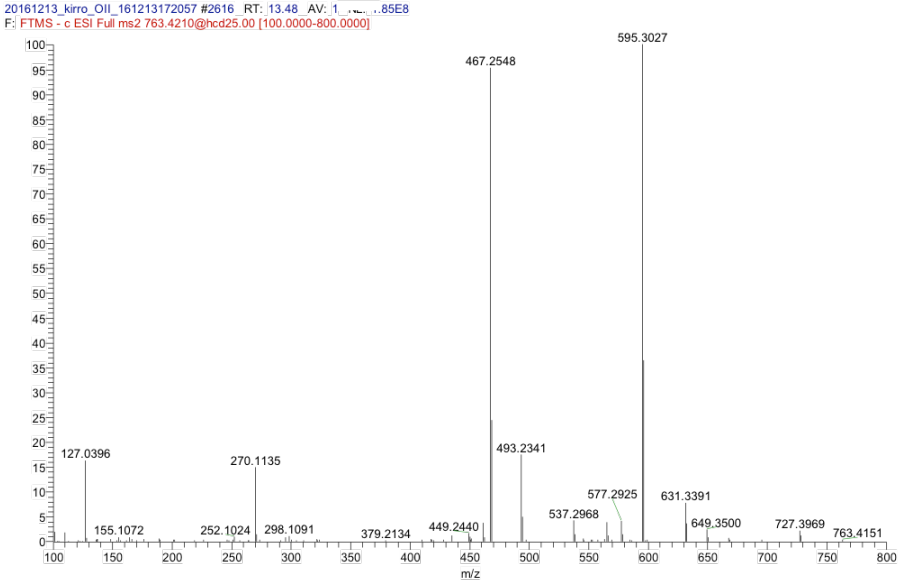
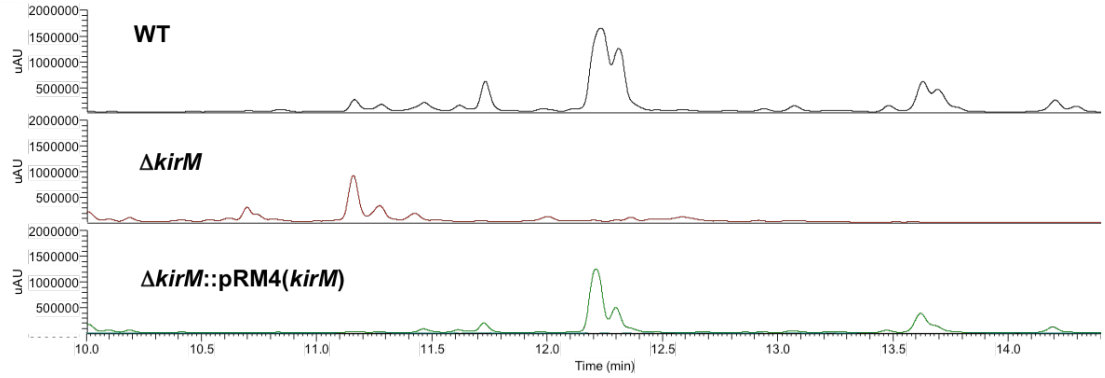

20161213_kirro_OII_161213172057 #2616 RT: 13.48 AV:
F: FTMS - c ESI Full ms2 763.4210@hcd25.00 [100.0000-800.0000]
595.3027
467.2548
493.2341
127.0396
270.1135
577.2925
631.3391
537.2968
649.3500
727.3969
763.4151
155.1072
252.1024
298.1091
379.2134
449.2440
m/z

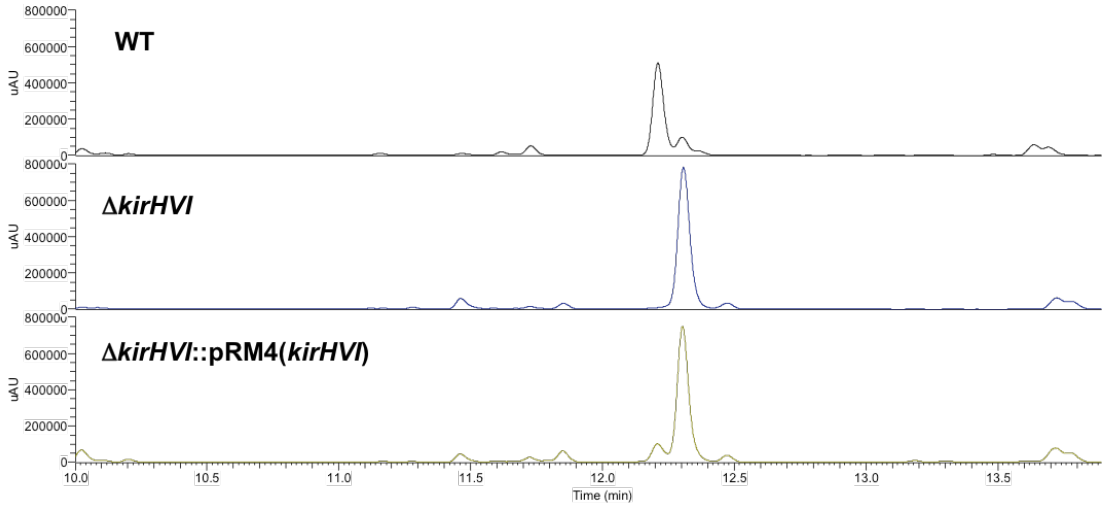
Figure S3-I – S3-VI: HPLC-ESI-UV-Vis profiles of the wild type (WT), Δ*kir* mutants, and complemented mutants

UV-Vis profiles of *Streptomyces collinus* Tü 365 (WT) compared to mutants **(I)** Δ*kirM*, **(II)** Δ*kirHVI*, **(III)** Δ*kirOI,* and, **(IV)** Δ*kirOII*, along with their respective complementations. The Δ*kirN* mutant and its single (*kirHVI*) and double (*kirN+HVI*) complementations are shown in **(V)** and **(VI)**, respectively.

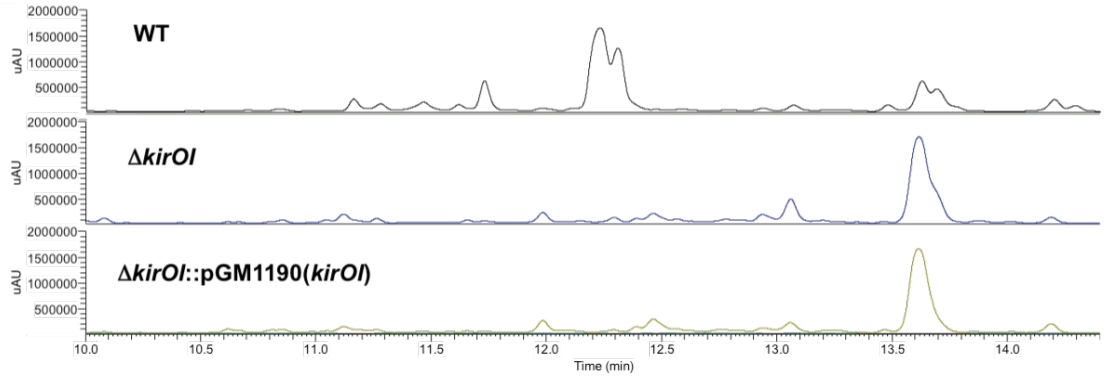
(I)



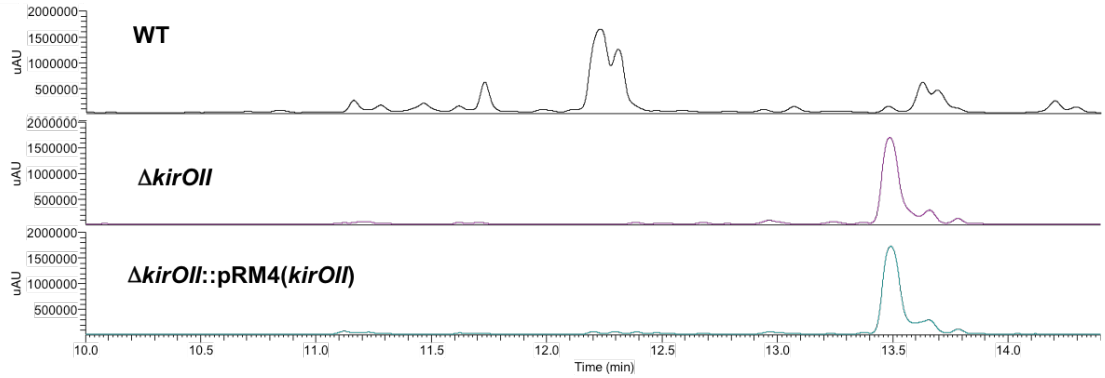
(II)



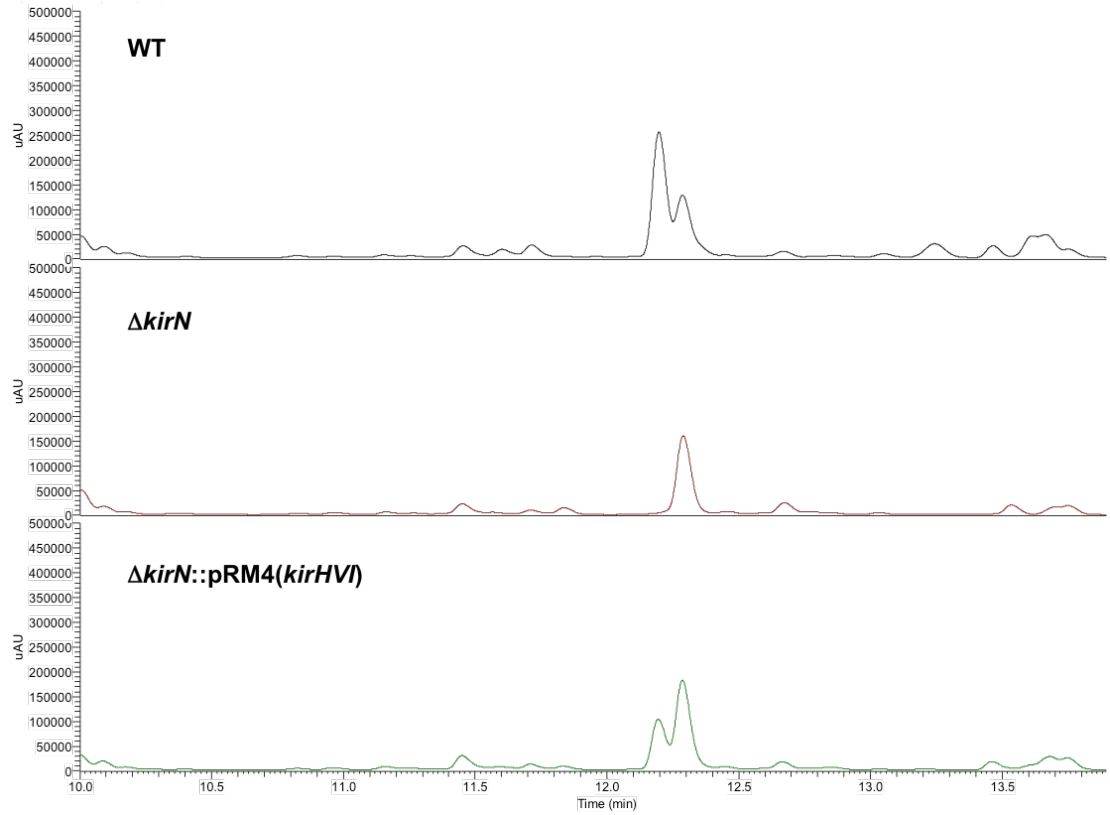
(III)


WT
ΔkirOI
ΔkirOI::pGM1190(kirOI)
uAU
2000000
1500000
1000000
500000
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
Time (min)

(IV)


WT
ΔkirOII
ΔkirOII::pRM4(kirOII)
uAU
2000000
1500000
1000000
500000
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
Time (min)

(V)


WT
ΔkirN
ΔkirN::pRM4(kirHVI)
uAU
Time (min)

(VI)

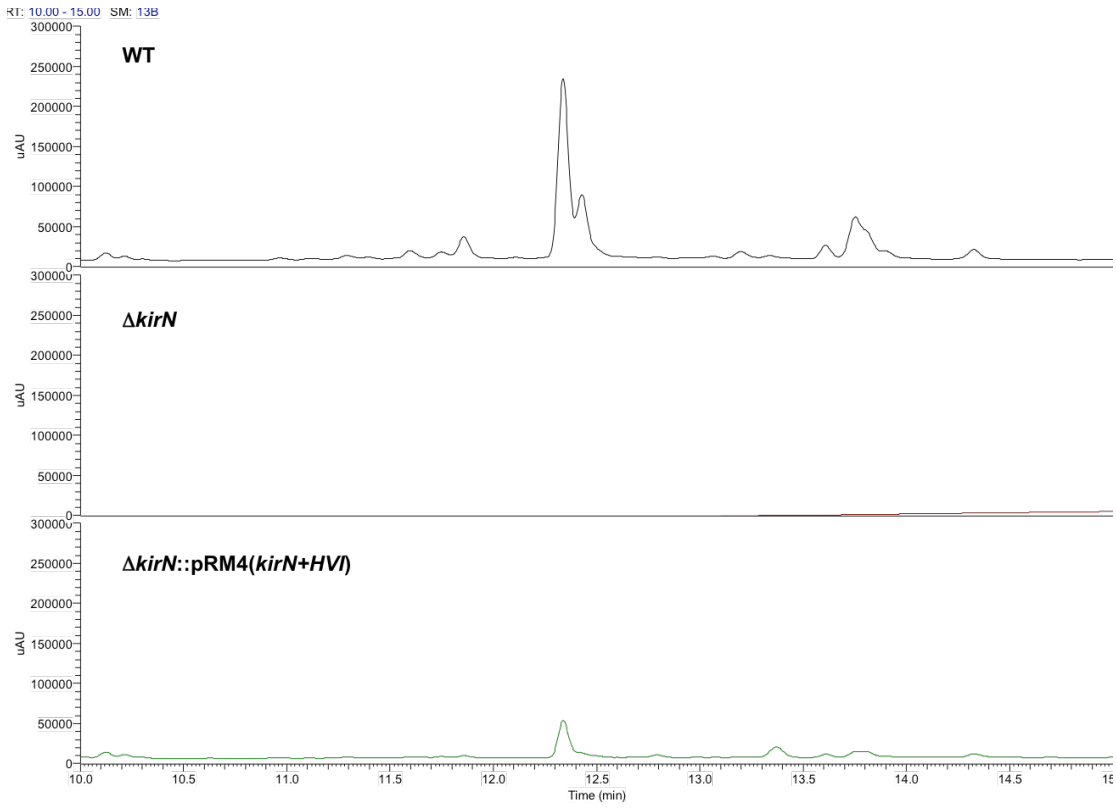
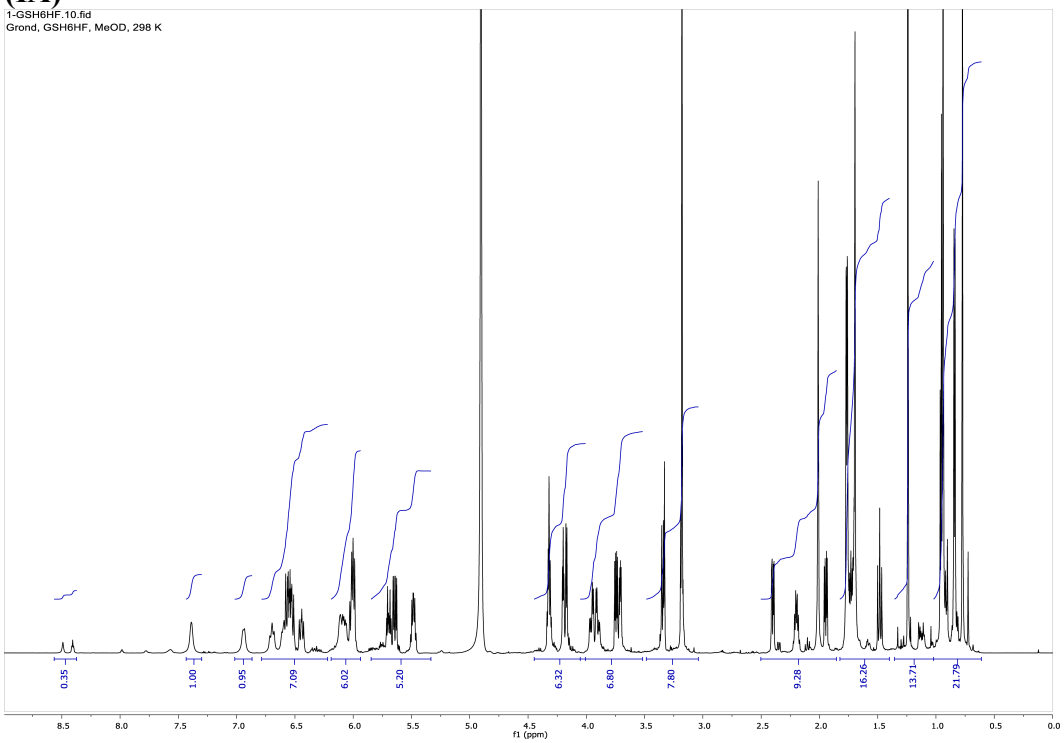

WT
ΔkirN
ΔkirN::pRM4(kirN+HVI)
uAU
Time (min)

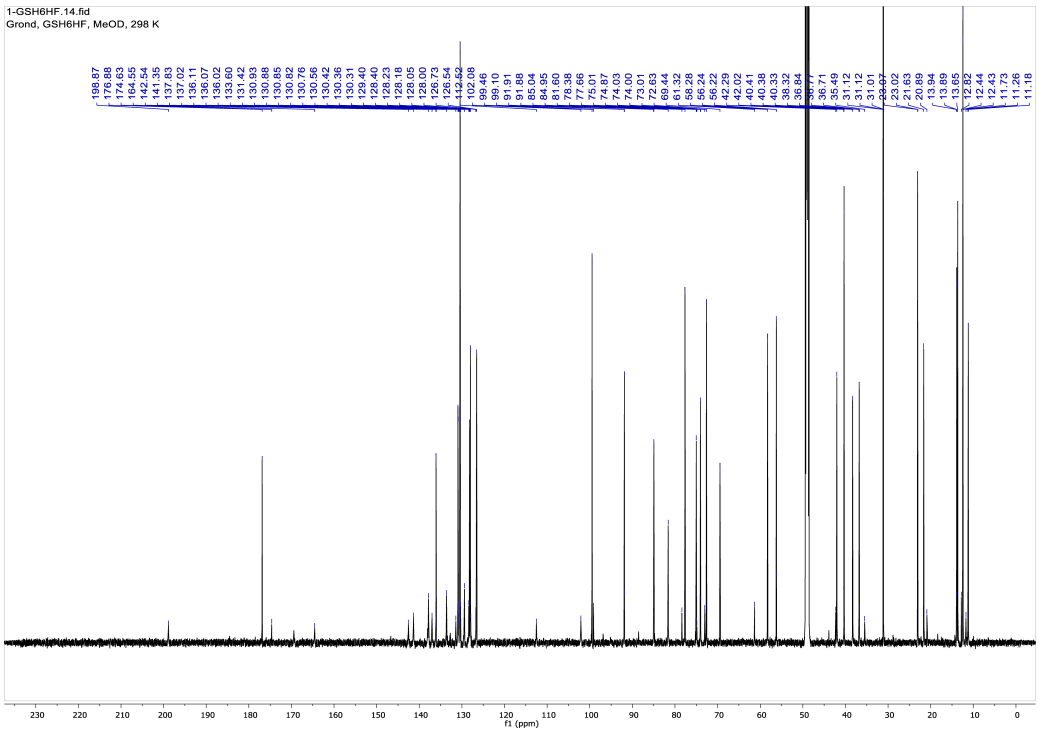
Figure S4-IA/B – S4-IIA/B: ^{1}HNMR and ^{13}CNMR data and assignment for kirromycin-HVI and kirromycin-OII

(**A**) ^{1}HNMR and (**B**) ^{13}CNMR of (**I**) kirromycin-HVI (**3**) and (II) kirromycin-OII (**5**) derivatives (CD_3OD, 298 K). Resonance frequency [1)] 600 resp. 150.8 MHz, [2)] 700 resp. 176.1 MHz).

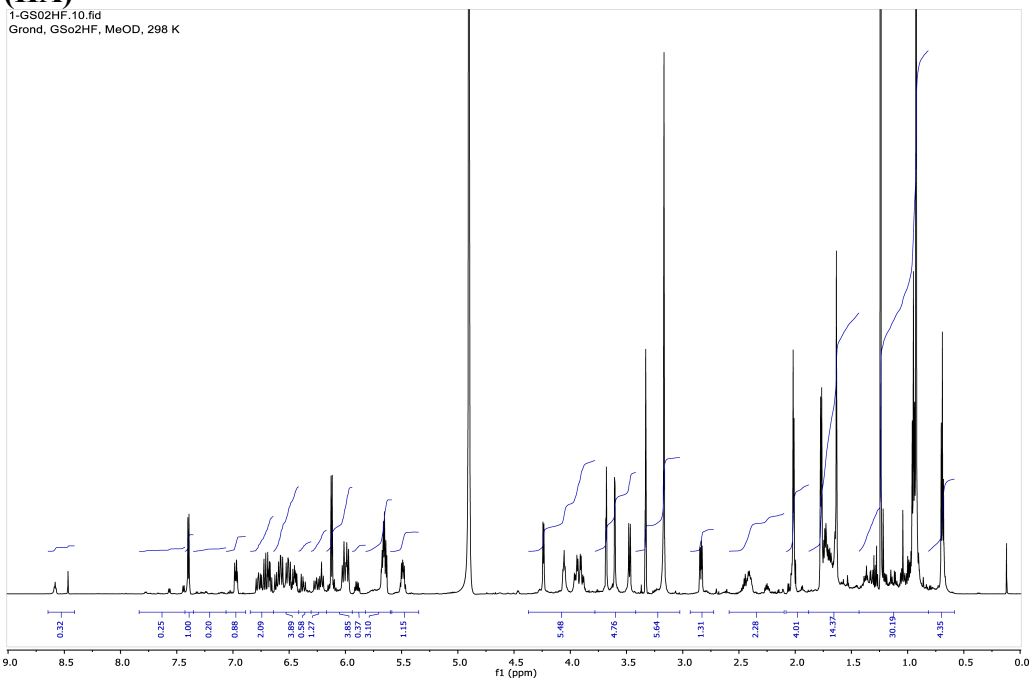
(IA)



(IB)



(IIA)


1-GS02HF.10.fid
Grond, GSo2HF, MeOD, 298 K
f1 (ppm)

(IIB)

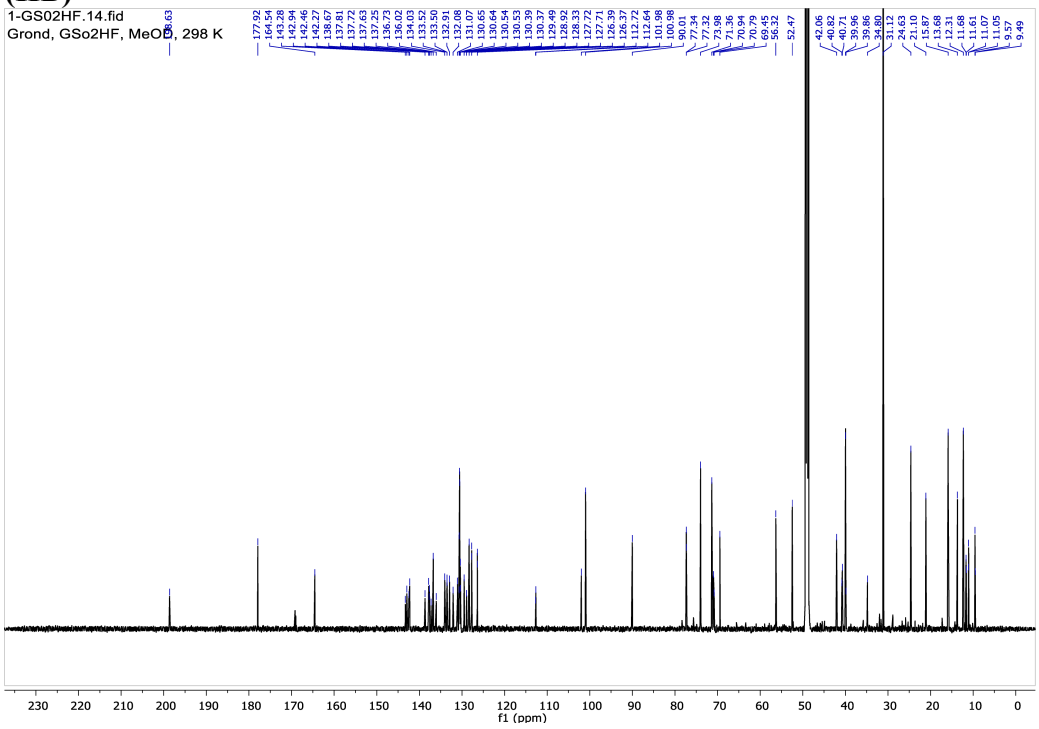

1-GS02HF.14.fid
Grond, GSo2HF, MeOD, 298 K
f1 (ppm)

Figure S5-I – S5-III: 2D NMR of kirromycin-OII (5) derivative

(I) Homonuclear correlation spectroscopy (COSY) spectrum established the partial structure for the disrupted furan part (**Fig. S5-A**) in the kirromycin-OII (**5**) derivative isolated from the Δ*kirOII* mutant. H17 (*δ* 4.05) exhibited correlations with both H18 (*δ* 1.69) and H16 (*δ* 2.42, *δ* 2.25). H16 exhibited further correlations to an olefinic proton H15 (*δ* 5.65), which confirmed the location of the extra double bond. **(III)** Heteronuclear multiple bond correlation (HMBC spectrum) showed correlations between H18, Me41, Me43 and C19 (*δ* 88.7) and corroborated the final structure. The red shifted UV-Vis spectrum for **5** could be explained by the elongation of the double bond chain. Additional 2D NMR spectra included: **(II)** (Heteronuclear single quantum correlation) HSQC.

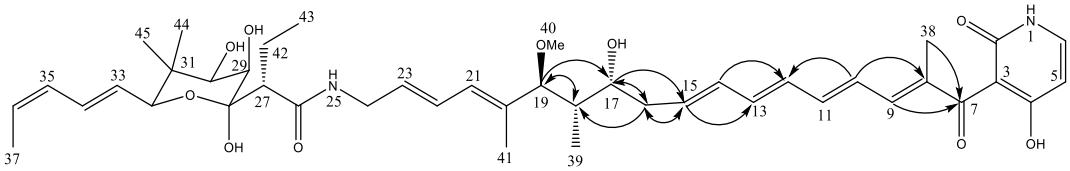
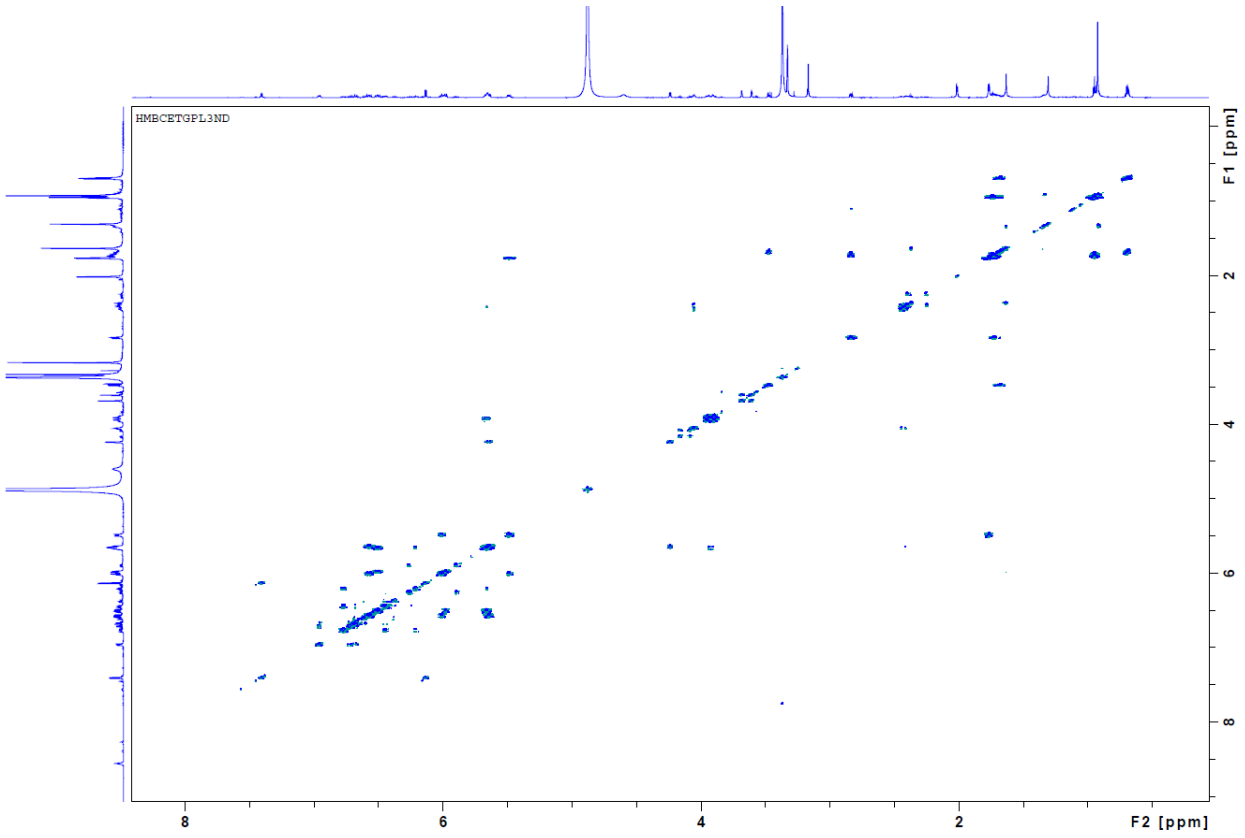
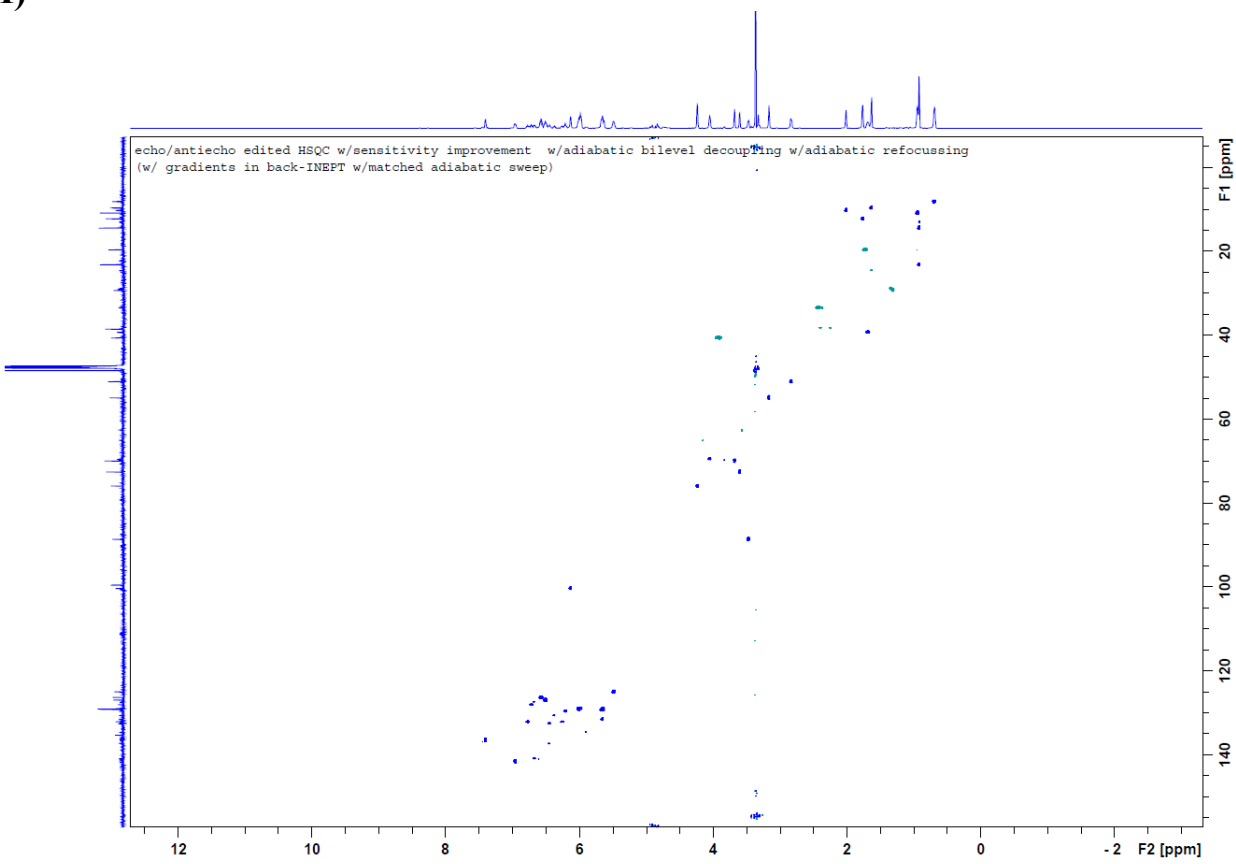


Figure S5-A Selected COSY (double arrowed) and HMBC (single arrowed) correlations for compound **5**.

(I)



(II)


echo/antiecho edited HSQC w/sensitivity improvement w/adiabatic bilevel decoupling w/adiabatic refocussing
(w/ gradients in back-INEPT w/matched adiabatic sweep)
F1 [ppm]
F2 [ppm]

(III)

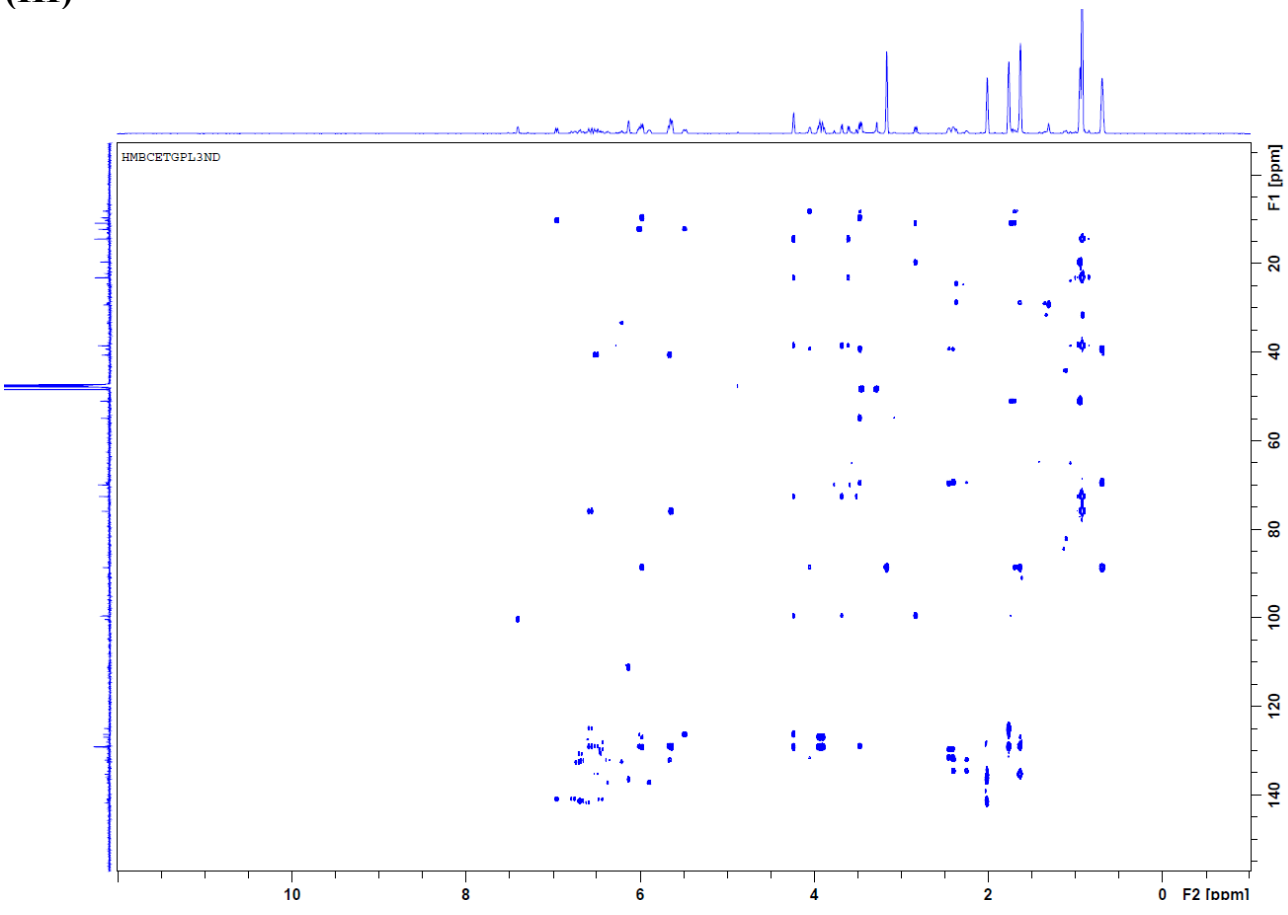
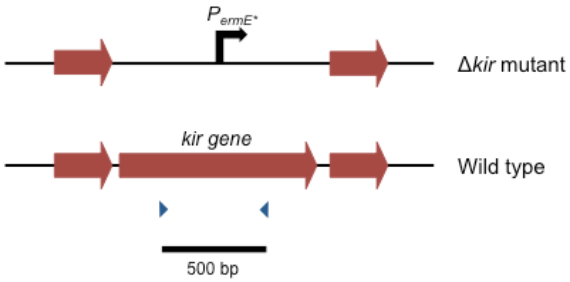

HMBCETGPL3ND
F1 [ppm]
F2 [ppm]

Figure S6-I – S6-II: Verification of Δ*kir* mutants by control PCRs

Control PCRs on genomic DNA (gDNA) of Δ*kirM,* Δ*kirHVI,* Δ*kirOI,* Δ*kirOII,* Δ*kirHIV,* Δ*kirHV,* and Δ*kirN* were carried out to confirm successful double crossover **(II)**. The use of primers (KP43 – KP56) targeting an internal 500 bp region of each gene in question **(I)** allowed for verification of correct mutants by the absence of bands (gene replaced by *ermE** promoter). *Streptomyces collinus* Tü 365 (positive) (**1**) and water (negative) (**2**) controls were included for each primer pair. GeneRuler™ 1 kb DNA ladder (Fisher Scientific) (**L**) was used as marker.

(I)



(II)

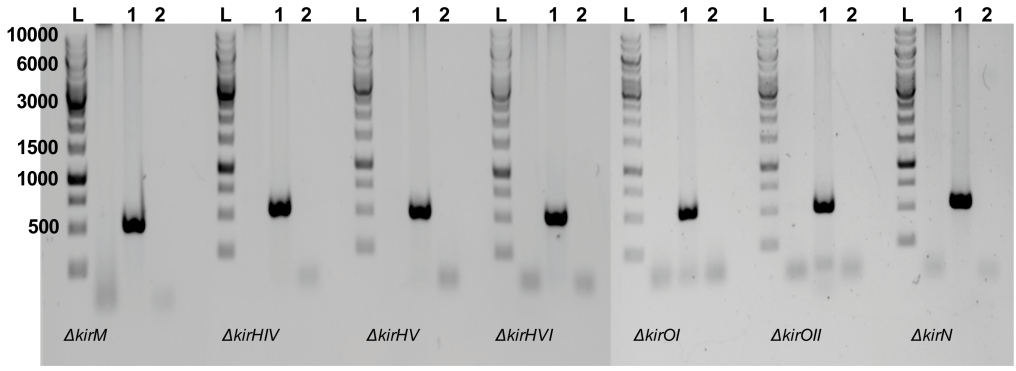
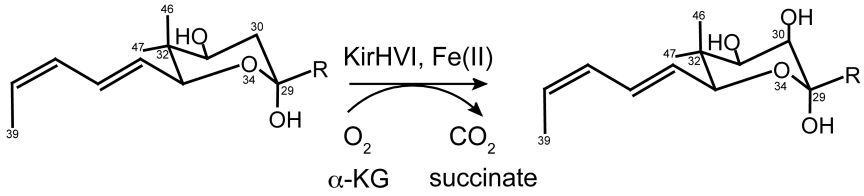


Figure S7-I – S7-III: Putative enzymatic reaction catalysed by tailoring enzymes KirHVI, KirOI, and KirOII

Genetic inactivation studies and NMR analysis enabled the prediction of the putative enzymatic reactions catalysed by the phytanoyl-CoA dioxygenase KirHVI (encoded by *kirHVI*) **(I)** and the two P450-dependent hydroxylases KirOI and KirOII (encoded by *kirOI* **(II)** and *kirOII* **(III)**.

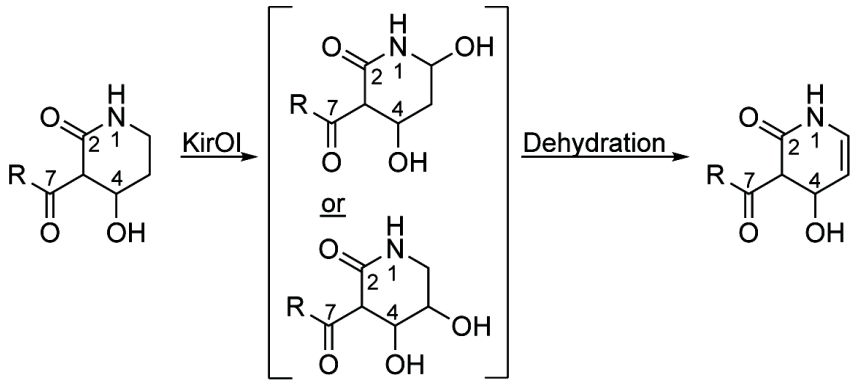
(I)

Based on similarity to the hydroxylase Fum3p in fumonisin B[9], KirHVI is postulated to induce α-ketoglutarate (α-KG)/Fe(II)-dependent hydroxylation at the C30 position in the sugar-like moiety in kirromycin.



(II)

The cytochrome P450-dependent hydroxylase KirOI catalyses hydroxylation of either C5 or C6 in the pyridone ring in kirromycin. The intermediate can then undergo dehydration, which might be facilitated by a yet uncharacterised enzyme or a *trans*-acting dehydrogenase (DH), to form the double bond between C5 and C6.

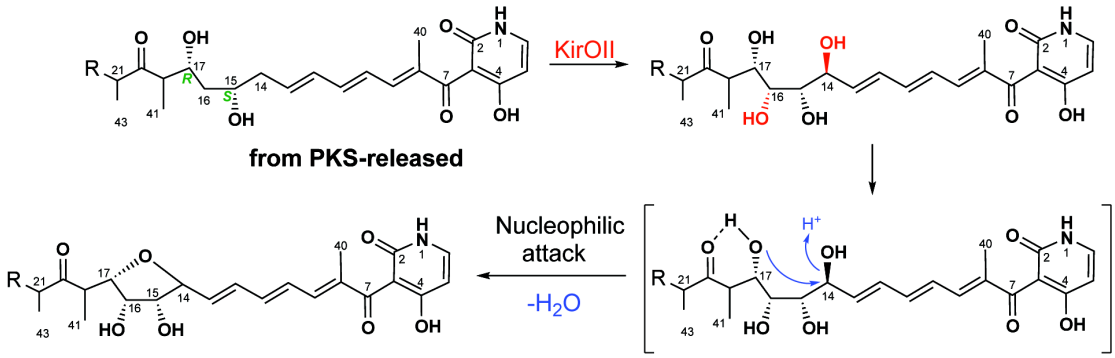


(III)

Our data suggests KirOII to be involved in formation of the tetrahydrofuran (THF) ring in the centre of the kirromycin molecule.

In case of kirromycin biosynthesis the PKS-derived precursor molecule of kirromycin could undergo dual CH_2-hydroxylation at C14 and C16, possibly catalysed by the cytochrome P450 monooxygenase KirOII. The hypothesis is strengthened from the sequence comparison of KirOII to AurH (31 % identity and 49 % similarity). The cytochrome P450 monooxygenase AurH has been found to be involved in formation of the THF ring in the antifungal compound aureothin produced by *Streptomyces thioluteus*[10,11] through the introduction of two C-O bonds. In a similar fashion, KirOII might catalyse the following step including a classical ring-closing dehydration reaction under formation of the THF ring moiety in kirromycin. This conversion is favoured in kirromycin because of increased nucleophilicity of C17-hydroxyl group as a result of the hydrogen bonding in a cyclic 6-membered ring including the C20 carbonyl group. From the Δ*kirOII* mutant, the tetraene product **4**, named 30-hydroxy-5,6-dehydro-1-*N*-demethyl-16-deoxy-kirrothricin, is isolated from the culture broth due to chemically favoured dehydration to the stable extended conjugated double bond system between C6 and C15.

While this is a plausible enzymatic reaction, the scheme currently lacks an additional enzyme responsible for catalysing the reduction of C=O to C-OH at the C20 position in kirromycin. Such an enzyme remains to be identified and characterized.



Supplementary references

1. MacNeil, D. J. *et al.* Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111,** 61–68 (1992).

2. Wolf, H. & Zähner, H. Stoffwechselprodukte von Mikroorganismen. *Arch. Mikrobiol.* **83,** 147–154 (1972).

3. Weber, T. *et al.* Molecular Analysis of the Kirromycin Biosynthetic Gene Cluster Revealed β-Alanine as Precursor of the Pyridone Moiety. *Chem. Biol.* **15,** 175–188 (2008).

4. Myronovskyi, M., Welle, E., Fedorenko, V. & Luzhetskyy, A. Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. *Appl. Environ. Microbiol.* **77,** 5370–5383 (2011).

5. Menges, R., Muth, G., Wohlleben, W. & Stegmann, E. The ABC transporter Tba of *Amycolatopsis balhimycina* is required for efficient export of the glycopeptide antibiotic balhimycin. *Appl. Microbiol. Biotechnol.* **77,** 125–134 (2007).

6. Muth, G., Wohlleben, W. & Puhler, A. The minimal replicon of the *Streptomyces ghanaensis* plasmid pSG5 identified by subcloning and Tn5 mutagenesis. *Mol. Gen. Genet.* **211,** 424–429 (1988).

7. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* **4,** 1020–1029 (2015).

8. Tong, Y., Robertsen, H. L., Blin, K., Weber, T. & Lee, S. Y. in *Methods in Molecular Biology* **1671,** 163–184 (2018).

9. Ding, Y., Bojja, R. S. & Du, L. Fum3p, a 2-Ketoglutarate-Dependent Dioxygenase Required for C-5 Hydroxylation of Fumonisins in *Fusarium verticillioides*. *Appl. Environ. Microbiol.* **70,** 1931–1934 (2004).

10. He, J. & Hertweck, C. Iteration as Programmed Event during Polyketide Assembly; Molecular Analysis of the Aureothin Biosynthesis Gene Cluster. *Chem. Biol.* **10,** 1225–1232 (2003).

11. Richter, M. E. A., Traitcheva, N., Knüpfer, U. & Hertweck, C. Sequential Asymmetric Polyketide Heterocyclization Catalyzed by a Single Cytochrome P450 Monooxygenase (AurH). *Angew. Chemie* **120,** 9004–9007 (2008).

Discussion & Perspectives

The overall aim of this thesis was the development and application of genome engineering tools for improved secondary metabolite production in streptomycetes, which are organisms known for their ability to produce a range of different clinically relevant compounds, including antibiotics.

Most antibiotics used in clinics today are based on synthetic or semi-synthetic natural product-derived compounds[94], however, with the increase in multidrug resistant pathogens, the need for new antibiotic classes are greater than ever. Fortunately, streptomycete genomes harbour several silent BGCs, which potentially could encode novel compound classes. With the discovery of these silent gene clusters and improvements of molecular tools for genome engineering, both small- and large-scale engineering of natural and heterologous *Streptomyces* hosts are becoming routine practice in many laboratories. A comprehensive review on the recent progress experienced within the field of streptomycetes engineering is presented in **manuscript I**[64] of this thesis.

In 2015, the CRISPR-Cas9 system was adapted for actinomycetes genome engineering[75–77]. Compared to the at that time existing molecular tools, the CRISPR-Cas9 system presented a relatively simple and efficient tool for manipulating actinomycetal DNA to refactor production of secondary metabolites. Reports of gene and gene cluster inactivations[75–77], gene silencing[77], and gene induction[80], using this system alone, have accelerated engineering efforts within the field.

A toolkit for CRISPR-Cas9 genome engineering of actinomycetes was recently developed by our group as a part of our compound discovery platform[77]. To encourage its application, our group extended the system into a more comprehensive toolkit, which has been included as **manuscript II**[95] in this thesis. The extended toolkit includes a software for sgRNA identification[96], a CRISPR-Cas9 toolbox for different genome engineering efforts, and an adapted uracil-specific excision reagent (USER)-cloning strategy and protocol for simple and efficient CRISPR plasmid construction. The CRISPR-Cas9 toolbox encompass different genome engineering tools, one of which relies on CRISPR-Cas9 plasmids for construction of random size libraries, which can be beneficial for gene function studies and for the generation of genome-minimised hosts[77]. Furthermore, the addition of HR templates in the

CRISPR-Cas9 vectors allows for scar-less gene deletions or gene insertions at defined sites in the host genome. Finally, the catalytically inactive dCas9 is included in the toolbox and represents a need tool for reversible gene repression. In this context, the development of a USER-cloning system for construction of the CRISPR-dCas9 vectors was included in **manuscript II** to facilitate future implementation of the system in an automated setting. In addition to the molecular tools, the CRISPy-web software was included. The software allows for computer-aided prediction of 20 bp recognition sequences for defined genes with as few mismatches, or off-targets, in the target genome as possible.

Further improvement to the USER-cloning system is reported in **manuscript III** of the thesis. While the genome engineering is based on the same principles as those in **manuscript II**, the major difference lies in the improved workflow for plasmid construction and its applicability in semi-automatic high throughput approaches. The new system includes the plasmids pUSER2-Cas9, pUSER2-dCas9, and pUSER2-Cas9/ScaLigD, which are all compatible with robotic equipment, in which the USER-assembly and clone verification can be automated with minimal operator interference. The assembled vectors can be directly transferred to *Streptomyces* or *Actinomyces* hosts.

The improved cloning system was verified in two manual proof-of-concept studies using pUSER-dCas9 for gene repression of core genes involved in actinorhodin (ACT) and undecylprodigiosin (RED) biosynthesis in *Streptomyces coelicolor* A3(2). While ACT and RED production could be easily evaluated due to the blue and red pigments, respectively, we realise that for future applications of the system, providing an additional link to local or global compound databases would improve dereplication of the complex sample mixtures substantially in terms of identifying novel compounds.

The feasibility of implementing a cloning system in a high-throughput platform has recently been explored for the construction of a library of transcription activator-like effector nucleases (TALENs) in an automated setting[97]. The synthesis platform was implemented on the programmable iBioFAB (Illinois Biofoundry for Advanced Biomanufacturing) system, which consists of single operational units capable of a range of molecular tasks, including PCR, DNA assembly, and transformation. Based on very limited operator interference, consisting of only reagents and consumables

supply, the automated workflow allowed for the high-throughput construction of a library of 192 TALEN constructs to be used for further testing.
The simplicity behind the USER assembly protocol allows for the same level of automation and would enable the fast and easy construction of a library of sgRNA-harbouring CRISPR constructs, requiring only limited operator interactions. The multiple applications of the CRISPR-Cas9 toolbox brings even more versatility to our system in that simple modifications, such as the exchange of destination vector or the inclusion of HR templates, give rise to entirely new constructs with different applications. In this sense, our platform does not restrict itself to single purposes, which further strengthens the incentive to implement the system on robotic equipment.

We believe that the combined efforts of **manuscript II** and **III** present a unique workflow with a high value for future engineering efforts, especially keeping in mind the ease of automation. In practise, our automated workflow would present itself as follows; First, the gifted strains entering the pipeline would undergo genome mining, using a tool such as antiSMASH[14,98], to predict and list gene targets for engineering. Some operator interference is expected at this stage in order to evaluate the output of the *in silico* prediction. Next, the chosen gene(s) would enter the CRISPy-web software for the automated prediction of 20 bp protospacers with as few off-targets as possible. The standardised USER-primer construction and ordering can be programmed as well. Supplying of the primers, templates, and linearised vectors is the last operator action needed for the plasmid construction. The assembled constructs can then be directly used for targeted gene inactivation or gene silencing.

The ability to generate several sgRNAs, both for single and multiple genes, would improve further downstream engineering efforts and could increase odds of a positive outcome in our compound discovery platform. On that note, it should be mentioned that in high-throughput settings, the robustness and consistency of the implemented protocol is of high importance in terms of reducing the screening required to obtain correct mutants. The current CRISPy-web software tool, implemented in our workflow, allows for prediction and evaluation of sgRNAs based on their number of off-targets. In the future, the prediction capabilities of our tool might be optimised to account for additional thermodynamic properties, which could improve efficacy of the sgRNA and with that the Cas9 cleavage[99,100]. As more studies are undertaken to

evaluate the specificity and efficiency of larger pools of sgRNAs, the applicability of the *in silico* prediction tools, such as the CRISPy-web, are expected to improve and expand. Being only recently implemented as a tool for genome engineering of streptomycetes, more systematic, large-scale studies on the individual components of the CRISPR-Cas9 system are expected in the future.

Streptomyces collinus Tü 365 is a known producer of the narrow-spectrum antibiotic kirromycin, which is produced from a hybrid PKS I/NRPS system[25]. Prior to our development of the optimised USER-CRISPR-Cas9 platform, the CRISPR-Cas9 editing tool[77] was tested for construction of a genetic variant of the original strain. The interest in using the CRISPR-Cas9 system for gene knockout in *S. collinus* Tü 365 was sparked from recent studies on the two discrete *trans*-AT enzymes KirCI and KirCII, which are involved in the supply of extender units to the assembly line of kirromycin[25–27]. The fact that KirCII can recognise and incorporate non-natural extender units, alternative to the ethylmalonyl-CoA, in the kirromycin precursor has led to the development of a system for polyketide bioderivatization[52]. Using this KirCII-ACP5 tool, the derivatives allyl- and propargyl-kirromycin was produced in the wild type strain *S. collinus* Tü 365. The major limitation of the system, however, remained in the lower production yields of the kirromycin derivatives compared to kirromycin, which presented a bottleneck for downstream processes, such as purification and pharmacokinetic studies. Since the pathway-specific CCR, encoded by *kirN*, is expected to be involved in supply of the ethylmalonyl-CoA, the first part of the thesis set out to investigate the role of this gene.

Unfortunately, our initial attempts to use the CRISPR-Cas9 system for gene inactivation of *kirN* resulted in unexpected genome deletions, in which both chromosomal arms of *S. collinus* Tü 365 were lost. As a consequence, both BGCs of kirromycin, found on the TIRs of the chromosome[49,50], were lost and kirromycin production abolished.

Chromosome instability in streptomycetes is not an uncommon phenomenon and has previously been linked with spontaneous mutations[88–90]. The mutations caused by the CRISPR-Cas9 or any other gene editing strategy are thus expected to give rise to instability, which might lead to aberrant reactions, such as those observed for the Δ*kirN* mutant. Furthermore, the natural instability at the telomeric regions on the chromosome of streptomycetes might provide additional explanation to the observed

phenotype of the Δ*kirN* mutant in that the BGC of kirromycin is found on the TIRs in the chromosome of *S. collinus* Tü 365. With our current data, it is not possible to exactly determine the series of events leading to the loss of both kirromycin BGCs in the Δ*kirN* mutant, but we suspect that the massive stress of the CRISPR-Cas9 mediated DSBs, that likely happen simultaneously on both TIRs, may lead to cyclisation of the chromosomes around the DSBs. This type of DSB-induced cyclisation has been reported previously for other streptomycetes, including *Streptomyces griseus*[101], *Streptomyces lividans* 66[102], and *Streptomyces ambofaciens*[103].

The growth of the Δ*kirN* mutant strain was unaffected compared to the wild type strain, which fits to previous reports of cyclised mutants growing with comparable efficiency to that of the parental strain[104]. A plausible explanation for this might be found in the unique organisation of the chromosomes of streptomycetes, in which genes monitoring core functions are located close to the *oriC*, located in the middle of the chromosome. However, in this respect it should be noted that circular chromosomes do not necessarily display higher stability than the linear[93,105].

As more laboratories start employing the CRISPR-Cas9 technology, reports describing this instability and possible explanations as to why it occurs might be something we see more of in the near future.

Due to the unexpected side effects of using the CRISPR-Cas9 system for genetic editing in *S. collinus* Tü 365, an alternative classical engineering method was selected, which makes use of a GusA reporter gene to indicate plasmid integration and loss. In continuation of the works of this thesis, our study on kirromycin biosynthesis was expanded to include not only *kirN*, but also genes putatively involved in tailoring reactions, as these were yet to be characterised. Six additional genes, *kirM, kirHVI, kirOI, kirOII*, *kirHIV,* and *kirHV* were included in the genetic study reported in **manuscript IV**.

Our studies on the genes involved in tailoring reactions in kirromycin biosynthesis allowed us to close most of the missing gaps in the biosynthetic pathway. For the genes *kirM, kirHVI, kirOI*, and *kirOII*, genetic inactivations led to the production of derivatives of kirromycin, which allowed for assigning the functions of the enzymes encoded by the genes. As a result, *kirM* was confirmed to encode an *O*-methyltransferase, responsible for installation of the methyl group at the O42 position

in kirromycin, and the dioxygenase *kirHVI* was found to be responsible for hydroxylation at the C30 position in the sugar-like moiety of the molecule. Furthermore, the cytochrome P450 hydroxylases, encoded by *kirOI* and *kirOII*, were found to be involved in double bond formation in the pyridone ring and for formation of the tetrahydrofuran (THF) ring, respectively. In respect to KirOII, our genetic and structural examinations of the kirromycin-OII derivative could open up for new studies on the role of hydroxylases in THF ring closure in the molecule. From our current data, the putative enzymatic function of KirOII is believed to be similar to that of the AurH, which is a cytochrome P450 monooxygenase responsible for installation of two C-O bonds during aureothin biosynthesis in *Streptomyces thioluteus* HKI-227[106,107]. To determine the exact enzymatic function of KirOII, the next step will be to carry out *in vivo* and/or *in vitro* enzymatic assays.

Using our alternative gene inactivation strategy, we also obtained a correct mutant in *kirN* with lowered kirromycin production compared to the wild type strain. This not only confirmed the role of *kirN* as a CCR, but also yielded a mutant suitable for the bioderivatization platform. We hypothesise that the reduction in ethylmalonyl-CoA in the Δ*kirN* mutant can give rise to a larger pool of precursors for the alternative malonyl-CoA extender units, such as propargylmalonyl-CoA. Using the KirCII-ACP5 tool with the Δ*kirN* mutant presents a future engineering scaffold to improve yields of propargyl-kirromycin or other derivatives. Propargyl-kirromycin is of special interest in that it can form the basis for further molecular derivatization based on Copper(I)-catalysed azide/alkyne cycloaddition (CuAAC), or "click" chemistry[52], to produce novel derivatives of kirromycin.

In addition to the derivatives generated with the Δ*kirN* scaffold, our studies on the four genes involved in tailoring reactions in kirromycin biosynthesis also provide us with a toolbox of four kirromycin derivatives. In the future, these new derivatives should be included in mode-of-action studies to test for potentially improved properties.

Recently, the Global Antibiotic Research & Development Partnership (GARDP), a joint initiative between the Drugs for Neglected Diseases *initiative* (DND*i*) and World Health Organisation (WHO), was launched to deliver new treatments for bacterial infections with emerging or present drug resistances[108]. One of the main parts of the initiative includes development of solutions to address the increase in sexually

transmitted infections (STIs) caused by the gram-negative pathogen *N. gonorrhoeae*[109]. Kirromycin has a reported MIC of 0.06 μg/mL[41] against *N. gonorrhoea*, but suffers from poor pharmacokinetic properties, which has thus far inhibited its used in the clinics. In this respect, the collection of new kirromycin derivatives, reported in our study, should be tested for improved pharmacokinetic properties. In addition, the Δ*kirN* mutant, with reduced kirromycin production, can be applied to the bioderivatization method to produce new kirromycin analogous with higher yields.

Conclusion

Engineering of secondary metabolism in streptomycetes has witnessed a great progress in recent years. As more whole genome sequence (WGS) projects are undertaken, our understanding of the potential of these gifted microorganisms for production of valuable compounds is increased. Furthermore, improvements to existing molecular cloning methods along with the development of entirely new ones are opening up for more rational and efficient engineering efforts.

This thesis contributes to the field of genome engineering of streptomycetes by the development of an improved USER-cloning methodology for our CRISPR-Cas9 toolbox, allowing for implementation of the system in a high-throughput semi-automatic setting. Lessons learned from the unintended side effects of using the CRISPR-Cas9 system for genetic inactivation in *S. collinus* Tü 365, has helped to uncover an area which should be subjected to more studies in the future. The fact that the molecular toolbox for genome engineering of streptomycetes continues to expand, allowed us to still study genes involved in tailoring reactions and ethylmalonyl-CoA supply in kirromycin biosynthesis in *S. collinus* Tü 365. This leaves us with a collection of mutants, which now can be tested for improved properties, especially in relation to treatment of infections caused by multidrug resistant *N. gonorrhoeae* isolates.

References

1. Chater, K. F. A Morphological and Genetic Mapping Study of White Colony Mutants of *Streptomyces coelicolor*. *J. Gen. Microbiol.* **72,** 9–28 (1972).

2. Wildermuth, H. & Hopwood, D. A. Septation During Sporulation in *Streptomyces coelicolor*. *J. Gen. Microbiol.* **60,** 51–59 (1970).

3. Claessen, D., Rozen, D. E., Kuipers, O. P., Søgaard-Andersen, L. & van Wezel, G. P. Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. *Nat. Rev. Microbiol.* **12,** 115–124 (2014).

4. Claessen, D. *et al.* The formation of the rodlet layer of streptomycetes is the result of the interplay between rodlins and chaplins. *Mol. Microbiol.* **53,** 433–443 (2004).

5. Flärdh, K. & Buttner, M. J. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.* **7,** 36–49 (2009).

6. Barka, E. A. *et al.* Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **80,** 1–43 (2016).

7. Williams, S. T. & Vickers, J. C. in *Biology of actinomycetes* (eds. Okami, Y., Beppu, T. & Ogawara, H.) 165–270 (Japan Scientific Societies Press, 1988).

8. Demain, A. L. & Sanchez, S. Microbial drug discovery: 80 years of progress. *J. Antibiot. (Tokyo).* **62,** 5–16 (2009).

9. Bérdy, J. Bioactive microbial metabolites. *J. Antibiot.* **58,** 1–26 (2005).

10. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417,** 141–147 (2002).

11. Zerikly, M. & Challis, G. L. Strategies for the Discovery of New Natural Products by Genome Mining. *ChemBioChem* **10,** 625–633 (2009).

12. Weber, T. *In silico* tools for the analysis of antibiotic biosynthetic pathways. *Int. J. Med. Microbiol.* **304,** 230–235 (2014).

13. Medema, M. H. & Fischbach, M. A. Computational approaches to natural product discovery. *Nat. Chem. Biol.* **11,** 639–648 (2015).

14. Blin, K. *et al.* antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45,** W36–W41 (2017).

15. Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx320

16. Weber, T. in *Encyclopedia of Industrial Biotechnology* (John Wiley & Sons, Inc., 2010). doi:10.1002/9780470054581.eib040

17. Weissman, K. J. in *Methods in Enzymology* **459,** 3–16 (2009).

18. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7,** 285–295 (2003).

19. Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chemie Int. Ed.* **48,** 4688–4716 (2009).

20. Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510,** 560–564 (2014).

21. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510,** 512–517 (2014).

22. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11,** 660–670 (2015).

23. Schwecke, T. *et al.* The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci.* **92,** 7839–7843 (1995).

24. Aparicio, J. F. *et al.* Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: Analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169,** 9–16 (1996).

25. Weber, T. *et al.* Molecular Analysis of the Kirromycin Biosynthetic Gene Cluster Revealed β-Alanine as Precursor of the Pyridone Moiety. *Chem. Biol.* **15,** 175–188 (2008).

26. Musiol, E. M. *et al.* Supramolecular Templating in Kirromycin Biosynthesis: The Acyltransferase KirCII Loads Ethylmalonyl-CoA Extender onto a Specific ACP of the *trans*-AT PKS. *Chem. Biol.* **18,** 438–444 (2011).

27. Musiol, E. M. *et al.* The AT(2) domain of KirCI loads malonyl extender units to the ACPs of the kirromycin PKS. *Chembiochem* **14,** 1343–1352 (2013).

28. Fernández-Moreno, M. A., Martínez, E., Boto, L., Hopwood, D. A. & Malpartida, F. Nucleotide-sequence and deduced functions of a set of cotranscribed genes of *Streptomyces coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin. *J. Biol. Chem.* **267,** 19278–19290 (1992).

29. Grimm, A., Madduri, K., Ali, A. & Hutchinson, C. R. Characterization of the *Streptomyces peucetius* ATCC 29050 genes encoding doxorubicin polyketide synthase. *Gene* **151,** 1–10 (1994).

30. Revill, W. P., Bibb, M. J. & Hopwood, D. A. Purification of a malonyltransferase from *Streptomyces coelicolor* A3(2) and analysis of its genetic determinant. *J. Bacteriol.* **177,** 3946–3952 (1995).

31. Arthur, C. J. *et al.* Self-Malonylation Is an Intrinsic Property of a Chemically Synthesized Type II Polyketide Synthase Acyl Carrier Protein. *Biochemistry* **44,** 15414–15421 (2005).

32. Pfeifer, V. *et al.* A Polyketide Synthase in Glycopeptide Biosynthesis. *J. Biol. Chem.* **276,**

38370–38377 (2001).

33. Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **20,** 79–110 (2003).

34. Funa, N. *et al.* A new pathway for polyketide synthesis in microorganisms. *Nature* **400,** 897–899 (1999).

35. Walsh, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **33,** 127–135 (2016).

36. Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529,** 235–238 (2016).

37. Gatto, G. J., McLoughlin, S. M., Kelleher, N. L. & Walsh, C. T. Elucidating the Substrate Specificity and Condensation Domain Activity of FkbP, the FK520 Pipecolate-Incorporating Enzyme. *Biochemistry* **44,** 5993–6002 (2005).

38. Du, L., Sánchez, C., Chen, M., Edwards, D. J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* **7,** 623–642 (2000).

39. Fischbach, M. A. & Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *ChemInform* **37,** 3468–3496 (2006).

40. Wolf, H. & Zähner, H. Stoffwechselprodukte von Mikroorganismen. *Arch. Mikrobiol.* **83,** 147–154 (1972).

41. Tavecchia, P. *et al.* Synthesis and biological evaluation of new fragments from kirromycin antibiotic. *J. Antibiot. (Tokyo).* **49,** 1249–1257 (1996).

42. Clough, B., Rangachari, K., Strath, M., Preiser, P. R. & Wilson, R. J. Antibiotic inhibitors of organellar protein synthesis in *Plasmodium falciparum*. *Protist* **150,** 189–195 (1999).

43. Leeds, J. A. *et al. In Vitro* and *In Vivo* Activities of Novel, Semisynthetic Thiopeptide Inhibitors of Bacterial Elongation Factor Tu. *Antimicrob. Agents Chemother.* **55,** 5277–5283 (2011).

44. Prezioso, S. M., Brown, N. E. & Goldberg, J. B. Elfamycins: inhibitors of elongation factor-Tu. *Mol. Microbiol.* **106,** 22–34 (2017).

45. Schmid, B., Anke, T. & Wolf, H. Action of pulvomycin and kirromycin on eukaryotic cells. *FEBS Lett.* **96,** 189–191 (1978).

46. Pingoud, A., Urbank, C., Wolf, H. & Maass, G. The Binding of Kirromycin to Elongation Factor Tu. Structural Alterations are Responsible for the Inhibitory Action. *Eur. J. Biochem.* **86,** 153–157 (1978).

47. Vogeley, L., Palm, G. J., Mesters, J. R. & Hilgenfeld, R. Conformational change of elongation factor Tu (EF-Tu) induced by antibiotic binding. Crystal structure of the complex between EF-Tu. GDP and aurodox. *J. Biol. Chem.* **276,** 17149–17155 (2001).

48. Carelli, J. D. *et al.* Ternatin and improved synthetic variants kill cancer cells by targeting the elongation factor-1A ternary complex. *Elife* **4,** e10222 (2015).

49. Rückert, C. *et al.* Complete genome sequence of the kirromycin producer *Streptomyces collinus* Tü 365 consisting of a linear chromosome and two linear plasmids. *J. Biotechnol.* **168,** 739–740 (2013).

50. Iftime, D. *et al.* Identification and activation of novel biosynthetic gene clusters by genome mining in the kirromycin producer *Streptomyces collinus* Tü 365. *J. Ind. Microbiol. Biotechnol.* **43,** 277–291 (2016).

51. Pavlidou, M. *et al.* The phosphopantetheinyl transferase KirP activates the ACP and PCP domains of the kirromycin NRPS/PKS of *Streptomyces collinus* Tü 365. *FEMS Microbiol. Lett.* **319,** 26–33 (2011).

52. Musiol-Kroll, E. M. *et al.* Polyketide Bioderivatization Using the Promiscuous Acyltransferase KirCII. *ACS Synth. Biol.* **6,** 421–427 (2017).

53. Gui, C. *et al.* Discovery of a New Family of Dieckmann Cyclases Essential to Tetramic Acid and Pyridone-Based Natural Products Biosynthesis. *Org. Lett.* **17,** 628–631 (2015).

54. Olsthoorn-Tieleman, L. N., Palstra, R.-J. T. S., van Wezel, G. P., Bibb, M. J. & Pleij, C. W. A. Elongation factor Tu3 (EF-Tu3) from the kirromycin producer *Streptomyces ramocissimus* is resistant to three classes of EF-Tu-specific inhibitors. *J. Bacteriol.* **189,** 3581–3590 (2007).

55. Bode, H. B., Bethe, B., Höfs, R. & Zeeck, A. Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem* **3,** 619–627 (2002).

56. Baltz, R. H. Genetic methods and strategies for secondary metabolite yield improvement in actinomycetes. *Antonie Van Leeuwenhoek* **79,** 251–259 (2001).

57. Stephanopoulos, G. Metabolic Fluxes and Metabolic Engineering. *Metab. Eng.* **1,** 1–11 (1999).

58. Ro, D.-K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440,** 940–943 (2006).

59. Murli, S., Kennedy, J., Dayem, L. C., Carney, J. R. & Kealey, J. T. Metabolic engineering of *Escherichia coli* for improved 6-deoxyerythronolide B production. *J. Ind. Microbiol. Biotechnol.* **30,** 500–509 (2003).

60. Hong, K.-K. & Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell. Mol. Life Sci.* **69,** 2671–2690 (2012).

61. Cho, C., Choi, S. Y., Luo, Z. W. & Lee, S. Y. Recent advances in microbial production of fuels and chemicals using tools and strategies of systems metabolic engineering. *Biotechnol. Adv.*

33, 1455–1466 (2015).

62. Fang, L., Zhang, G. & Pfeifer, B. A. in *Functional Metagenomics: Tools and Applications* (eds. Charles, T. C., Liles, M. R. & Sessitsch, A.) 45–63 (Springer International Publishing, 2017). doi:10.1007/978-3-319-61510-3_3

63. Baltz, R. H. Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes. *J. Ind. Microbiol. Biotechnol.* **43,** 343–370 (2016).

64. Robertsen, H. L., Weber, T., Kim, H. U. & Lee, S. Y. Towards systems metabolic engineering of streptomycetes for secondary metabolites production. *Biotechnol. J.* 1700465 (2017). doi:10.1002/biot.201700465

65. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science (80-.).* **315,** 1709–1712 (2007).

66. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annu. Rev. Genet.* **45,** 273–297 (2011).

67. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482,** 331–338 (2012).

68. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337,** 816–821 (2012).

69. Makarova, K. S., Aravind, L., Wolf, Y. I. & Koonin, E. V. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct* **6,** 38 (2011).

70. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).

71. Nishimasu, H. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156,** 935–949 (2014).

72. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157,** 1262–1278 (2014).

73. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32,** 347–355 (2014).

74. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).

75. Cobb, R. E., Wang, Y. & Zhao, H. High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **4,** 723–728 (2015).

76. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-

mediated genome editing in *Streptomyces*. *Acta Biochim. Biophys. Sin. (Shanghai).* **47,** 231–243 (2015).

77. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* **4,** 1020–1029 (2015).

78. Qin, Z. *et al.* Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African *Tetraponera* plant-ants. *Chem. Sci.* **8,** 3218–3227 (2017).

79. Wolf, T. *et al.* Targeted genome editing in the rare actinomycete *Actinoplanes* sp. SE50/110 by using the CRISPR/Cas9 System. *J. Biotechnol.* **231,** 122–128 (2016).

80. Zhang, M. M. *et al.* CRISPR–Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters. *Nat. Chem. Biol.* **13,** 607–609 (2017).

81. Jia, H. *et al.* Development of a CRISPR/Cas9-mediated gene-editing tool in *Streptomyces rimosus*. *Microbiology* **163,** 1148–1155 (2017).

82. Wilson, M. C. & Moore, B. S. Beyond ethylmalonyl-CoA: The functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* **29,** 72–86 (2012).

83. Deatherage, D. E. & Barrick, J. E. in *Methods in Molecular Biology* **1151,** 165–188 (2014).

84. Zakrzewska-Czerwińska, J. & Schrempf, H. Characterization of an autonomously replicating region from the *Streptomyces lividans* chromosome. *J. Bacteriol.* **174,** 2688–2693 (1992).

85. Musialowski, M. S. Functional evidence that the principal DNA replication origin of the *Streptomyces coelicolor* chromosome is close to the *dnaA-gyrB* region. *J. Bacteriol.* **176,** 5123–5125 (1994).

86. Choulet, F. *et al.* Evolution of the Terminal Regions of the *Streptomyces* Linear Chromosome. *Mol. Biol. Evol.* **23,** 2361–2369 (2006).

87. Perry, J. A. & Wright, G. D. The antibiotic resistance 'mobilome': searching for the link between environment and clinic. *Front. Microbiol.* **4,** 138 (2013).

88. Leblond, P. & Decaris, B. New insights into the genetic instability of *Streptomyces*. *FEMS Microbiol. Lett.* **123,** 225–232 (1994).

89. Dharmalingam, K. & Cullum, J. Genetic instability in *Streptomyces*. *J. Biosci.* **21,** 433–444 (1996).

90. Altenbuchner, J. & Cullum, J. DNA amplification and an unstable arginine gene in *Streptomyces lividans* 66. *MGG Mol. Gen. Genet.* **195,** 134–138 (1984).

91. Chen, C. W. *et al.* The Linear Chromosomes of *Streptomyces*: Structure and Dynamics. *Actinomycetologica* **8,** 103–112 (1994).

92. Volff, J. N. & Altenbuchner, J. Influence of disruption of the *recA* gene on genetic instability and genome rearrangement in *Streptomyces lividans*. *J. Bacteriol.* **179,** 2440–2445 (1997).

93. Volff, J. N., Viell, P. & Altenbuchner, J. Artificial circularization of the chromosome with concomitant deletion of its terminal inverted repeats enhances genetic instability and genome rearrangement in *Streptomyces lividans*. *Mol. Gen. Genet. MGG* **253,** 753–760 (1997).

94. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Last 25 Years. *J. Nat. Prod.* **70,** 461–477 (2007).

95. Tong, Y., Robertsen, H. L., Blin, K., Weber, T. & Lee, S. Y. in *Methods in Molecular Biology* **1671,** 163–184 (2018).

96. Blin, K. CRISPy-web: An online resource to design sgRNAs for CRISPR applications. *Synth. Syst. Biotechnol.* **1,** 118–121 (2016).

97. Chao, R. *et al.* Fully Automated One-Step Synthesis of Single-Transcript TALEN Pairs Using a Biological Foundry. *ACS Synth. Biol.* **6,** 678–685 (2017).

98. Medema, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39,** W339-46 (2011).

99. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).* **346,** 1258096–1258096 (2014).

100. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34,** 184–191 (2016).

101. Kameoka, D. *et al.* Analysis of fusion junctions of circularized chromosomes in *Streptomyces griseus*. *J. Bacteriol.* **181,** 5711–5717 (1999).

102. Redenbach, M. *et al.* The *Streptomyces lividans* 66 chromosome contains a 1 MB deletogenic region flanked by two amplifiable regions. *MGG Mol. Gen. Genet.* **241,** 255–262 (1993).

103. Leblond, P. *et al.* The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. *Mol. Microbiol.* **19,** 261–271 (1996).

104. Chen, C. W., Huang, C.-H., Lee, H.-H., Tsai, H.-H. & Kirby, R. Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. *Trends Genet.* **18,** 522–529 (2002).

105. Fischer, G., Decaris, B. & Leblond, P. Occurrence of deletions, associated with genetic instability in *Streptomyces ambofaciens*, is independent of the linearity of the chromosomal DNA. *J. Bacteriol.* **179,** 4553–4558 (1997).

106. He, J. & Hertweck, C. Iteration as Programmed Event during Polyketide Assembly; Molecular Analysis of the Aureothin Biosynthesis Gene Cluster. *Chem. Biol.* **10,** 1225–1232 (2003).

107. Richter, M. E. A., Traitcheva, N., Knüpfer, U. & Hertweck, C. Sequential Asymmetric Polyketide Heterocyclization Catalyzed by a Single Cytochrome P450 Monooxygenase

(AurH). *Angew. Chemie* **120,** 9004–9007 (2008).

108. Alirol, E. *et al.* Multidrug-resistant gonorrhea: A research and development roadmap to discover new medicines. *PLOS Med.* **14,** e1002366 (2017).

109. Unemo, M. & Shafer, W. M. Antimicrobial Resistance in *Neisseria gonorrhoeae* in the 21st Century: Past, Evolution, and Future. *Clin. Microbiol. Rev.* **27,** 587–613 (2014).

Appendix

1. Abbreviations

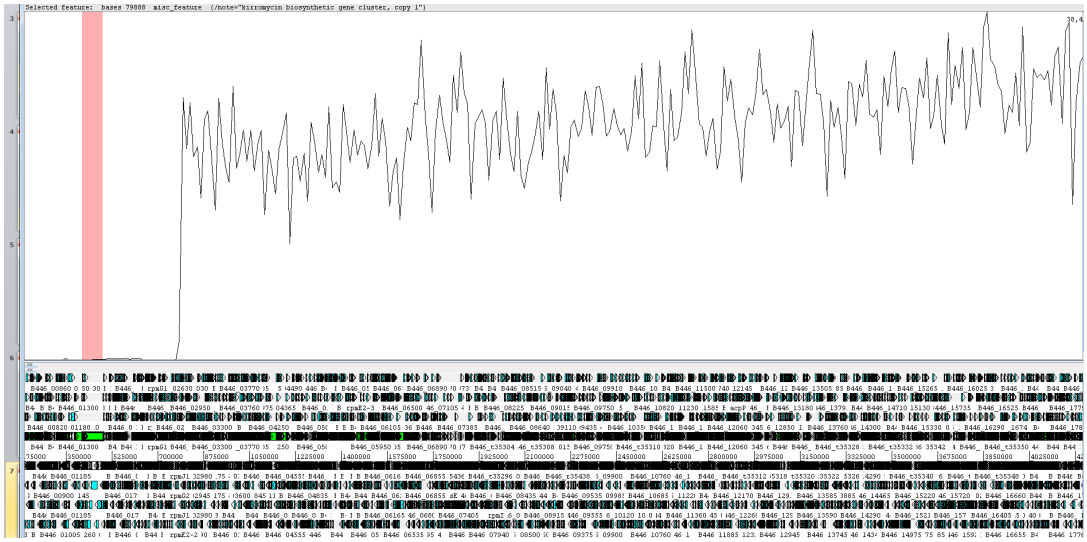
A	adenylation
aa	amino acid
ACP	acyl carrier protein
ACT	actinorhodin
ARO	aromatase
AT	acyltransferase
BGC	biosynthetic gene cluster
C	condensation
Cas	CRISPR-associated
CCR	crotonyl-CoA reductase/carboxylase
CID	collision-induced dissociation
CmR	chloramphenicol resistance marker
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
Cryo-EM	cryogenic electron microscopy
CuAAC	Copper-dependent azide-alkyne cycloaddition
CYC	cyclase
dCas9	catalytically inactive Cas9
DH	dehydratase
DSB	double stranded break
dU	uracil base
E	epimerisation
EF	elongation factor
ER	enoyl reductase
FA	fatty acid
GNAT	Gcn5-related N-acetyltransferase
GOI	gene of interest
HDR	homology directed repair
HR	homologous recombination
HRMS	high-resolution mass spectrometry
IC_{50}	half maximal inhibitory concentration
KR	ketoreductase
KS	ketosynthase
LB	Luria broth
MCAT	malonyl-CoA:ACP transferase
MIC	minimal inhibitory concentration
MS/MS	tandem mass spectrometry
MT	methyltransferase
NHEJ	non-homologous end joining

NMR	nuclear magnetic resonance
NRPS	non-ribosomal peptide synthetase
nt	nucleotide
ORF	open reading frame
OSMAC	One strain-Many compounds
PAM	protospacer associated motif
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
PKS	Polyketide synthase
PoC	proof-of-concept
ppm	parts per million
PPTase	Sfp-type phosphopantetheinyl transferase
pre-crRNA	precursor CRISPR RNA
PS	pyran synthase
RED	undecylprodigiosin
S.O.C.	Super Optimal broth with Catabolite repression
SFM	soy flour-mannitol medium
sgRNA	single guide RNA
T	thiolation
Ta	annealing temperature
TE	thioesterase
TIR	terminal inverted repeat
tracrRNA	*trans*-activating crRNA
TSB	tryptic soy broth
UHPLC	ultra high performance liquid chromatography
USER	uracil-specific excision reagent
WGS	whole genome sequence
WT	wild type
X-Gluc	5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid

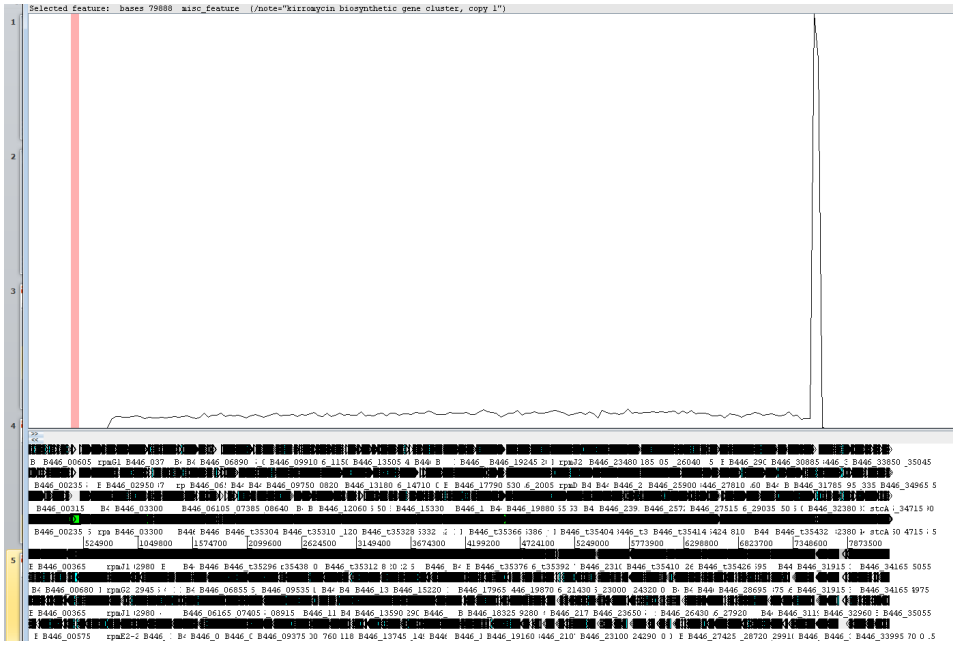
2. Artemis evaluation of the Δ*kirN* mutant of *Streptomyces collinus* Tü 365

An installable version of the free genome browser Artemis (http://www.sanger.ac.uk/science/tools/artemis) was used to align the reads obtained from Illumina MiSeq sequencing of the Δ*kirN* mutant against the reference genome of *Streptomyces collinus* Tü 365. The two screenshots below display the zoom-in (**2AI**) and zoom-out (**2AII**) of the bam-sorted alignment. As seen in both (**2AI**) and (**2AII**), the loss of the both the chromosomal arms in the mutant are evident from the missing read coverage. The zoom-in (**2AI**) is on the left arm of the chromosome of Δ*kirN* mutant.

(2AI)



(2AII)



3. Breseq analysis of the Δ*kirN* mutant of *Streptomyces collinus* Tü 365

The open-source computational pipeline Breseq has been developed for identification and annotation of genetic differences in mutants based on the mapping of data, obtained from next-generation sequencing technology, against a high-quality reference genome[83]. Using this tool, the Illumina MiSeq reads of the Δ*kirN* mutant was mapped against the reference genome of *Streptomyces collinus* Tü 365. The output data was a HTML file containing a list of all mutations found in the mutant. A modified version of the list, in which only the large deletions are listed, is presented in **Fig. 3AI** below. As seen from the list, both arms of the chromosome of the Δ*kirN* mutant were lost. The list include start and end positions of the two large deletions in addition to their sizes, which are of 787,795 and 630,478 bp. Genes involved in kirromycin biosynthesis has been underlined (in red) in the figure.

(3AI)

Unassigned missing coverage evidence
seq id
start
end
size
+-reads
reads+-
gene
description
610 genes
CP006259
1
787795
787795
NA [0]
[0] 1
B446_00005–B446_03105]
463 genes
CP006259
764244
8272925
6004781
[0] NA
[B446_32920]-B446_35285]